Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Stanford University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
HERSCHEL W. LEIBOWITZ, *International Business Machines Corporation, Bethesda, Maryland*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

The Effects of Differential Visual Stimulation after Induction of Visual Aftereffects: H. L. PICK, JR., M. HETHERINGTON, AND R. BELKNAPP 425

Effects of Probable Outcome Information on Two-Choice Learning: R. C. NIES 430

Color Coding and Visual Search: S. L. SMITH 434

Resistance to Extinction When Partial Reinforcement Is Followed by Regular Reinforcement: H. M. JENKINS 441

Partial Reinforcement, Continuous Reinforcement, and Reinforcement Shift Effects: S. H. HULSE 451

Inferential Behavior in Children as a Function of Age and Subgoal Constancy: T. S. KENDLER AND H. H. KENDLER 460

The Semantic Mediation of Evaluative Meaning: F. J. DI VESTA AND D. O. STOVER 467

Reversal and Nonreversal Shifts in Concept Formation Using Consistent and Inconsistent Responses: M. HARROW AND A. M. BUCHWALD 476

The Serial Position Effect of Free Recall: B. B. MURDOCK, JR. 482

The Scaling of Subjective Roughness and Smoothness: S. S. STEVENS AND J. R. HARRIS 489

An Evaluation of the Activationist Hypothesis of Human Vigilance: J. A. ADAMS AND L. R. BOULTER 495

The Relative Efficiency of Several Training Methods as a Function of Transfer Task Complexity: G. E. BRIGGS AND J. C. NAYLOR 505

Recruitment, Latency, Magnitude, and Amplitude of the GSR as a Function of Interstimulus Interval: W. F. PROKASY, J. T. FAWCETT, AND J. F. HALL 513

(Continued on inside back cover)

American Psychological Association

Vol. 64, No. 5 November 1962

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.* When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1962 the publication cost is $30 per printed page.

All **regular** articles, **early publication** articles, **supplementary reports,** and **replication reports** should be submitted to the Editor-elect: David A. Grant: Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR, Managing Editor
ELIZABETH S. REED, Advertising Manager
RUTH P. JACOLINZER, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 64, NO. 5 NOVEMBER 1962

THE EFFECTS OF DIFFERENTIAL VISUAL STIMULATION AFTER INDUCTION OF VISUAL AFTEREFFECTS[1]

HERBERT L. PICK, JR., MAVIS HETHERINGTON, AND ROLAND BELKNAPP

University of Wisconsin

The relationship between Gibson's negative aftereffect and the Köhler-Wallach figural aftereffect has been the subject of controversy.

Gibson (1933) found that a curved line, when perceived for a period of time, becomes phenomenally less curved than it was originally, and that after such an inspection period an objectively straight line appears curved in the opposite direction from the inspection curve. These negative aftereffects were attributed to a process similar to sensory adaptation in which curvature perceptions when long continued tend to approach the norm of a straight line. Gibson regards the straight line as a neutral from which other lines deviate. Thus a straight line will serve as an anchoring point in the perception of curved lines. A frequent condition of the environment tends to become a norm of the phenomenal world and new stimuli are perceived in relation to it.

Köhler and Wallach (1944) in their studies of figural aftereffects postulated an electrical field process in the visual cortex which satiates the cortex in the immediate area of the cortical representation of the inspection figure. This satiation results in increased resistance to further stimulation in this area and to displacement of the cortical representation to neighboring regions upon subsequent stimulation. Köhler and Wallach proposed that Gibson's curved line effect could be adequately explained by their satiation theory, if the test line were considered to be displaced from the satiation area of the inspection curve.

Although Osgood and Heyer (1952) suggest an alternative mechanism to explain figural aftereffects, both their theory and the Köhler-Wallach theory are based upon the same sort of physiological satiation by prolonged inspection of contours.

Several recent attempts have been made to differentiate experimentally the Gibson negative aftereffect from the Köhler-Wallach figural aftereffect. Sagora and Oyama (1957) in their survey of studies of figural aftereffects in Japan cited considerable evidence to indicate that the curved line effect could not be explained by the Köhler-Wallach theory. Bergman and Gibson (1959), using a slanted textured surface as a stimulus, demonstrated that one type of negative aftereffect could not be explained on the basis of satiation and contour displacement. Gibson (1959a) has argued on logical grounds that the two kinds of aftereffect cannot possibly be the same, and has recently (Gibson, 1959b, pp. 489–491) restated the normalization hypothesis.

Such evidence as that reported by Sagora and Oyama (1957) and Berg-

[1] This research was supported in part by the Research Committee of the Graduate School of the University of Wisconsin with funds provided by the Wisconsin Alumni Research Foundation. Authors are indebted to J. J. Gibson for a critical and helpful reading of the manuscript.

man and Gibson (1959) suggest that different processes are involved in the two types of aftereffect. Further investigation seemed warranted to elucidate the nature of these differences and to clarify the theoretical bases of the phenomena.

The Köhler-Wallach theory suggests a gradual dissipation with time of the differential satiation in the visual cortex. If the dissipation is spontaneous, and if satiation explains both contour-displacement and curvature-straightening, it could be predicted that with homogeneous stimulation (e.g., a "ganzfeld") interpolated between the viewing of the inspection figure and of the test figure, both types of visual aftereffects should decrease. The Gibson theory, on the other hand, suggests that the normalization of a curved line would not dissipate during an afterperiod of homogeneous stimulation since the aftereffect represents simply a readaptation to the conditions of stimulation prevailing before the experiment. If the effects of homogeneous stimulation on the two types of visual aftereffects differ, the distinctiveness of the two phenomena will have been demonstrated and the adequacy of either theory to explain such differential results can then be evaluated.

METHOD

In order to test the above predictions three different postinspection visual experiences were introduced following the induction of a Köhler-Wallach aftereffect and a Gibson curved line aftereffect, and their influence upon the aftereffects was then measured.

Subjects.—The *S*s for the Köhler-Wallach figural aftereffect were 36 summer school students registered in introductory psychology courses; *S*s for the Gibson negative aftereffect were 30 summer school students enrolled in introductory psychology or education courses.

Apparatus.—The apparatus for the Köhler-Wallach effect was a modified Dodge type tachistoscope in which the inspection (I) figure was shown binocularly at a distance of 58 cm. from the eye. The test (T) figures were exposed at an equal distance. In both cases the visual field was 19.7 cm. square. The I figure consisted of a circle 5.5 cm. in diameter which was placed to the left of a fixation point in the center of the visual field. The width of the line defining the circumference of the circle was 2 mm. Each T figure was drawn on white cardboard and consisted of two circles placed on opposite sides and equally distant from a fixation point. The width of the line in the T circles was .75 mm. In half the test trials the circle on the left was 7.6 cm. in diameter and would tend to be expanded outward to produce an aftereffect of increased size. These circles were always concentric with the original I figure. When the left circle was 7.6 cm. in diameter, the circle on the right was 7.6, 7.8, 8.0, 8.2, or 8.4 cm. in diameter. In the other half of the test trials the circle on the left was 4.0 cm. in diameter and would tend to be compressed inward to produce an aftereffect of decreased size. When the circle on the left of the T figure was 4.0 cm. in diameter the circle on the right was 3.6, 3.7, 3.8, 3.9, or 4.0 cm. in diameter. In both cases the size of the right circle was varied randomly from trial to trial. The use of two different size circles on the left was necessitated by the possibility of response bias occurring if the figural aftereffects always occurred in the same direction. A measure of the magnitude of the figural aftereffect could be determined by noting size judgments of circles on the right in relation to circles on the left. The method of obtaining this measure is discussed in the section on scoring of performance.

The Gibson visual aftereffect was produced by having *S*s fixate an I figure consisting of a black curved line 3 mm. in width and convex to the left drawn on white cardboard. In order to eliminate straight reference lines, the I line was presented through a circular aperture 30 cm. from the eye. The diameter of the aperture was 25.5 cm. The curved line extended the full length of the aperture and was bowed 2 cm. at its center. A series of five T lines was constructed and presented in the same manner. The T lines, 3 mm. in thickness, were also bowed to the left by 0.0, 0.1, 0.2, 0.3, and 0.5 cm. The T lines were presented in a random order. The magnitude of the aftereffect could be determined by noting which of these lines were perceived as straight.

The brightness of the larger surfaces (walls,

table-top) to which *S* was exposed under conditions of normal stimulation varied from 4.0 to 4.4 ft-c as measured by a Macbeth illuminometer. The goggles for the ganzfeld condition passed approximately 25% of the illumination. The brightness of the I and T fields for the Köhler-Wallach figural aftereffects was 2.7 ft-c. The brightness of the I and T field for the Gibson aftereffects was 3.2 ft-c.

Procedure.—Hammer (1949), in the only published study of the dissipation of figural aftereffects, found that they decreased to zero in 150 sec. Pilot studies, using the conditions outlined above, indicated that with 5-min. I periods and subsequent normal visual conditions the Köhler-Wallach aftereffect diminished to zero in approximately 30 min.; the Gibson effect in 15 min. These time intervals were consequently used for the postinduction visual exposure periods in the respective experiments. Each experiment consisted of five steps in the following order: an initial control test with the T figures, a 5-min. fixation of the I figure, an immediate test of the magnitude of the aftereffect, a postinspection period in one of three visual conditions, and a final test of the magnitude of the aftereffect.

In the Gibson procedure the control test consisted in making judgments of the direction of curvature of the five curved test figures convex to the left plus two curved figures convex to the right. The latter curved lines were included to accustom *S* to perceive lines curved in both directions. These lines were presented in random order. This procedure required approximately 20 sec. depending on the speed of judgment of *S*. During the 5-min. I period which followed the control test, *S* was instructed to run his eyes slowly up and down the middle portion of the I line. The immediate test of the magnitude of the aftereffect was the same as the control test except for the omission of the curves convex to the right. One-third of the *S*s then received each of the following postinspection conditions for 15 min.: Cond. I, normal stimulation, *S* looked around the room; Cond. II, homogeneous light stimulation, *S*'s vision was limited to a homogeneous ganzfeld, produced by goggles which covered each eye with a concave section of a translucent ping pong ball; Cond. III, homogeneous lack of stimulation, *S* had a black blindfold over open eyes. A final test of the magnitude of the aftereffect was administered as before.

In the Köhler-Wallach procedure the initial control test consisted of viewing the 10 test cards with two black circles on them. Cards were presented in random order at 4-sec. intervals. The *S* was instructed to fixate on the cross between the circles and report which of the circles was larger or if they were equal in size. During the 5-min. I period *S* was directed to fixate on the cross on the I figure. The subsequent test utilized the same procedure and stimuli as the control test. One-third of the *S*s were then exposed for 30 min. to each of the three postinspection conditions described above and finally the third test of the magnitude of the figural aftereffect was administered in the same manner as the previous ones.

Scoring of performance.—The scoring procedure for the Köhler-Wallach aftereffect was as follows: An arbitrary scoring system assigned numerical values to the size of the figural aftereffect. If the 7.6-cm. circle on the right was judged to be equal to the 7.6-cm. circle on the left (standard), a score of zero was assigned; if it was judged smaller (this being evidence of a figural aftereffect), a score of +1 was assigned. If the 7.8-cm. circle on the right was judged smaller than the standard, a score of +2 was assigned. Thus the weights were increased by integral units as the magnitude of figural aftereffect increased. An identical procedure was used with the 4.0-cm. standard circle, but since the effect of the inspection was to make the 4.0-cm. standard appear smaller, evidence of the visual aftereffect was a tendency of the circle on the right to appear larger.

A similar procedure was used to score the magnitude of the aftereffect of the Gibson type.

Results and Discussion

With the scoring procedures a value was computed for each *S* for the control test, the initial aftereffect test, and the final aftereffect test. These were computed by totaling the number of points obtained by *S* on all the test stimuli. The average values obtained and their *SD*s are shown in Table 1. In each experiment the values obtained for each of the three groups could be compared. The differences between the magnitude of the aftereffect (immediately after inspection) and the final magnitude (after the interpolated visual experience) were analyzed with each *S* serving as

TABLE 1

AVERAGE MAGNITUDES AND *SD*s OF AFTEREFFECT SCORES

Test	Köhler-Wallach						Gibson					
	Normal		Blindfold		Ganzfeld		Normal		Blindfold		Ganzfeld	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Control	2.8	3.5	2.6	3.8	2.0	2.1	0.1	0.3	0.1	0.7	0.2	0.4
Initial aftereffect	6.6	4.2	7.2	5.4	5.7	3.3	3.0	2.0	2.8	1.9	4.4	2.2
Final aftereffect	1.8	4.1	5.6	4.2	8.1	3.0	0.9	1.0	0.6	0.5	0.6	0.5

his own baseline. Negative differences occurred where the final aftereffect test resulted in a lower score than the immediate postinspection test, i.e., a decrease of the aftereffect over time. Conversely, positive differences indicated an increase of aftereffect over time.

The results shown in Table 1 suggest that with the Köhler-Wallach experiment the ganzfeld condition serves to enhance the aftereffect, and although the blindfold condition does not enhance the aftereffect it does reduce the amount of decay. The greatest decrease in aftereffect occurred with the normal vision group. With the Gibson aftereffect the greatest decay occurred under the ganzfeld condition. These findings are verified when the data are subjected to analysis of variance.

With the Köhler-Wallach effect the three groups are significantly different from each other ($P < .01$) and the ganzfeld group shows a significant enhancement effect ($P < .05$), i.e., an increase in magnitude of aftereffect over the 30-min. period. A significant difference ($P < .025$) between groups found with the Gibson aftereffect is apparently due to the fact that the ganzfeld group had somewhat inflated values in the initial aftereffect test. In any case no significant differences in final level of aftereffect occurred with the Gibson aftereffect, i.e., the aftereffect had vanished completely under all three visual conditions.

Since conditions of homogeneous stimulation effect these two visual aftereffects in radically different ways, it is strongly implied that the two phenomena are basically different.

The results are, in general, opposite to what might be expected if each theory predicted the outcome for its own type of aftereffect. The inference from the Köhler-Wallach theory was that a diminution of the aftereffect would occur following conditions of homogeneous stimulation. However, the Köhler-Wallach aftereffect was enhanced by the ganzfeld situation. Gibson's normalization theory would predict a maintenance of the curved line aftereffect under homogeneous stimulation but instead this effect had essentially vanished.

The nature of the perceptual system which might account for the paradoxical enhancement or the maintenance of the Köhler-Wallach effect warrants further investigation. It may be related to the higher susceptibility to figural aftereffects found after a period of sensory deprivation (Doane, Manatoo, Heron, & Scott, 1959). In the present study, if the ganzfeld and blindfold conditions were considered conditions of sensory deprivation it would suggest that stimulation just prior to such deprivation has a particularly strong effect. If the homogeneous conditions are related to sensory deprivation the differences obtained under the two conditions require further investigation.

Summary

The effects of three conditions of post-stimulation on a Köhler-Wallach figural after-effect and a Gibson negative aftereffect were investigated. Condition I was normal stimulation, obtained by looking around the room; Cond. II was homogeneous lack of stimulation, obtained by wearing a black blindfold; Cond. III was homogeneous light stimulation obtained by exposure to a "ganzfeld." The Gibson aftereffect decreased normally under all three conditions. The ganzfeld enhanced the Köhler-Wallach aftereffect, the blindfold retarded the decrease in the aftereffect, and looking around the room permitted the normal disappearance of the figural aftereffect.

References

Bergman, R., & Gibson, J. J. The negative after-effect of the perception of surface slanted in the third dimension. *Amer. J. Psychol.*, 1959, **72,** 364–374.

Doane, B. K., Manatoo, W., Heron, W., & Scott, T. H. Changes in perceptual function after isolation. *Canad. J. Psychol.*, 1959, **13,** 210–219.

Gibson, J. J. Adaptation, after-effect, and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16,** 1–31.

Gibson, J. J. After-effects: Figural and negative. *Contemp. Psychol.*, 1959, **4,** 294–295. (a)

Gibson, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. I. New York: McGraw-Hill, 1959. Pp. 456–501. (b)

Hammer, E. R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, **62,** 337–354.

Köhler, W., & Wallach, H. Figural after-effects. *Proc. Amer. Phil. Soc.*, 1944, **88,** 269–357.

Osgood, C. E., & Heyer, A. W., Jr. A new interpretation of figural-effects. *Psychol. Rev.*, 1952, **59,** 98–118.

Sagora, M., & Oyama, T. Experimental studies on figural after-effects in Japan. *Psychol. Bull.*, 1957, **54,** 327–338.

(Received October 9, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 430–433

EFFECTS OF PROBABLE OUTCOME INFORMATION ON TWO-CHOICE LEARNING

RICHARD C. NIES[1]

University of California, Los Angeles

The impact of statistical learning theory (Bush & Mosteller, 1955; Estes, 1950) has focused attention upon probability learning in the two-choice situation (Goodnow, 1958; Humphreys, 1939). Although success is maximized by consistent choice of the more likely alternative, *Ss* typically approximate the actual proportions of the reinforcement schedule (Hake, 1955; Jenkins & Stanley, 1951). However, when the task is presented in a gambling context (i.e., total correct choices are maximized), *Ss* tend to predict the more frequent event at a significantly higher proportion of the trials (Goodnow, 1955; Siegel & Goldstein, 1959). The present study was designed to further explore the conditions which maximize success by manipulating information about the probability of the events. Although responses in two-choice situations appear to be relatively independent of experimental instructions (Anderson & Grant, 1957; Neimark & Shuford, 1959) and performance information (Das, 1961), Koehler (1961) was able to predict mean terminal response rates by varying instructions which dealt with how *Ss* should consider the nonreinforced trials in a two-choice contingent partial reinforcement situation. In view of this finding, the stability of behavior to the manipulative effects of differential information in such situations is to be questioned. It was therefore hypothesized that specific information about the probabilities of two alternative outcomes would produce a shift toward consistent selection of the more likely event.

[1] The author expresses his appreciation to Allen Parducci and Norman H. Anderson for their invaluable guidance in the design and analysis of this experiment.

METHOD

Apparatus.—A box containing 100 marbles, 70 of one color and 30 of another color, was mounted so that the marbles could be mixed by turning a crank protruding from one end. A trough at the other end received 1 marble whenever the box was tilted forward, the marble rolling back into the box when the apparatus was returned to its normal position.

Subjects.—Eight experimental and four control sessions were held, with groups of 16 *Ss* drawn for each session from the course in introductory psychology at the University of California, Los Angeles.

Procedure.—The following instructions were read aloud to all *Ss* in the experimental groups:

> This experiment is designed to study human guessing habits. In this box are a number of blue and red marbles (shown briefly). For each trial, I will shake the box thoroughly and allow one marble to roll into the trough (demonstrated). While I am shaking the box I will count to three (demonstrated), and by the time I say "three," you write down on the response sheet, which will be given you shortly, whether you think a blue marble or a red marble will roll out. After you have written your choice, I will allow a marble to roll out and you will then be told its color. Your task is to get as many correct predictions as you can.

To assure the 70-30 ratio of blue to red marbles, a secret sheet was prepared which maintained this schedule randomly in blocks of 50 (35 blue marbles to 15 red marbles) over 250 trials. The random sequences were chosen with no constraints from a table of random numbers according to their probabilities. Although an assistant read the color of the marble for each trial from this prepared

schedule, *Ss* were led to believe that he was reporting the color of the marble that actually rolled out. This procedure insured that the different groups were exposed to identical sequences.

The *Ss* in each experimental session were randomly divided into four equal subgroups, differentiated with respect to the additional information printed at the top of their response sheets. The No-Information group had no additional information. The Pattern group had the following sentence added: "It has been found in experiments of this kind that the marbles roll out in definite patterns." The Ratio group had the statement: "There are 100 marbles in the box: 70 are blue and 30 are red." And the Ratio-Explanation group was informed that: "There are 100 marbles in the box: 70 are blue and 30 are red. This means that the chances of a blue marble rolling out are 7 out of 10, and the chances of a red marble rolling out are 3 out of 10. Since each marble that rolls out is returned to the box, these odds will be the same for each trial. Furthermore, since the box is shaken thoroughly each time, there can be no fixed pattern in which the marbles roll out."

For all conditions, *S* was to indicate his prediction by recording either an "R" or a "B" on each trial.

To control for color preferences and sequence peculiarities, half the experimental *Ss* were run with colors reversed (i.e., 70 reds to 30 blues), and half were run using a different 70-30 sequence. To control for unique group differences, each of these four experimental combinations was replicated with a second group. The result was a 2 × 2 × 4 factorial design (Color × Sequence × Information) replicated with a second set of four experimental groups.

Four control groups (two groups for each reinforcement sequences under the .70-red condition only) were also run to provide a baseline against which the effects of the marble box could be assessed. They were read the following instructions:

> This experiment is designed to study human guessing habits. For each trial, I will call out one of two colors—blue or red. Before I announce the color, I will count to three (demonstrated). While I am counting, you write down on the response sheet, which will be given you shortly, whether you think I will call out blue or red. After you have written your choice, I will then announce the color. Your task is to get as many correct predictions as you can.

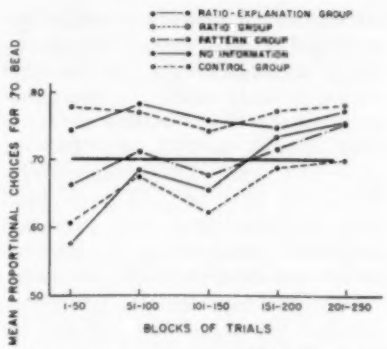

FIG. 1. Mean proportional choices for the .70 marble by trial blocks under various levels of information.

RESULTS AND DISCUSSION

Figure 1 shows the mean proportion of choices for the more likely alternative over each successive block of 50 trials. Separate analyses of variance were performed for each block of 50 trials, using the mean proportions as raw scores. Table 1 summarizes the analysis for the first 50 trials, showing the significant effect of Information and also of the interaction between Information and Sequence. This interaction was found only in the first

TABLE 1

ANALYSIS OF VARIANCE FOR TRIAL BLOCKS 1–50 FOR EXPERIMENTAL GROUPS

Source	*df*	*MS*	*F*
Instructions (I)	3	639.80	14.55**
Color (C)	1	54.82	1.44
Sequence (S)	1	2.32	.06
I × C	3	5.55	.12
I × S	3	156.18	3.53*
C × S	1	19.33	.51
I × C × S	3	40.88	.92
Groups within treatments[a]	4	38.10	.86
Pooled Groups × Instructions	12	51.91	1.17
Ss within groups[b]	96	44.28	

[a] Error term for C, S, and C × S.
[b] Error terms for I, I × C, I × S, I × C × S, Groups within treatments, and Pooled Groups × I.
* Significant at .05 level.
** Significant at .001 level.

block of 50 trials, so it was not judged as critical in modifying the interpretation of the main effects. In the other blocks, the only significant effect that appeared was that for Information in Trial Blocks 51–100 and 101–150 ($F = 3.64$ and 3.61, respectively; $df = 3/96$; $P < .05$). The different levels of information thus produced significant differences in betting behavior over the first three-trial blocks.

Since the experimental conditions revealed no significant effects for Trials 201–250, the scores for this block were pooled (using the .70-red condition only) and tested against the control scores for the same block. The difference was highly significant ($F = 16.33$; $df = 1/100$; $P < .001$).

These analyses support the following conclusions: (*a*) The information about probable event outcomes affect *Ss*' responses during the early blocks of trials (up to 150 trials), but shows no significant effect in later blocks. These early differences appear to be relatively independent of color preferences, sequence peculiarities, and group differences. (*b*) While the control groups (no marbles) just reach the probability matching level on the last block of trials, the experimental groups (marbles) exceed this level for all the trial blocks. Relevant to this finding is the proposal by Flood (1954) and Rubinstein (1959) that the awareness of randomness in a two-choice probabilistic outcome situation tends to elevate *S*'s response predictions, the impossibility of a complete solution increasing *S*'s caution against betting on the less likely alternative. Randomness was made explicit in the present study by the use of a box in which marbles were thoroughly mixed in *Ss*' presence before each prediction was made. In the control groups with no marbles, *Ss* displayed typical probability matching behavior.

While consistent with the interpretation made by Flood and Rubinstein, the present experiment advances their argument by demonstrating an early facilitation of optimal betting through the introduction of specific information relevant to the probable outcome of the events. While the presence of the marble box can account for the higher response level of the experimental groups, presumably through its contribution to the random appearance of the events, the perception of randomness was hastened by the critical information supplied to Ratio and Ratio-Explanation groups. When information is effective in determining behavior in a two-choice, uncertain outcome situation, its effectiveness may be based on its contribution to the perception of randomness. The fact that the instructions used by Anderson and Grant (1957) and Neimark and Shuford (1959) did not elevate *Ss*' response levels may result from a failure to make explicit the randomness and impossibility of a complete solution.

Although the experimental conditions elevated the proportion of correct anticipations, only 4 *Ss* (all from the Ratio-Explanation group) learned to bet 100% on the more likely alternative. Two suggestions for this relative lack of optimal betting find support in the present research: (*a*) The *Ss* seem surprisingly expectant of patterns and systems when there is any challenge of a problem to be solved (Goodnow, 1958). This is consistent with the responses to a questionnaire given at the conclusion of the present experiment where 75 of the 128 experimental *Ss* reported looking for a pattern whereas only 32 had been told (falsely) that there was one. Also, as shown in Fig. 1, the betting curve for the Pattern group is fairly close to that of the No-Information group. Moreover, while patterns may be expected from the problem solving nature of the task, certain characteristics of the data support it. In contrast to the method of constrained randomization used by Edwards (1961; Lindman & Edwards, 1961) and Nicks (1959), the method used in this research involved an unrestrained randomization, resulting in a distribution of lengths of homogeneous outcome runs which included too many short runs and not enough long runs. Thus there were

patterns in the outcome sequence which allowed for the gambler's fallacy. The point to be made here is that this reduction from optimal betting is not "irrational" in terms of the *Ss*' set to expect patterns. (*b*) Furthermore, the *Ss* prefer the occasional success of guessing the unlikely alternative to the monotony of repetitiously predicting the more certain event where no tangible inducements are offered to maximize their correct predictions (Goodnow, 1955; Siegel & Goldstein, 1959). Thus, there is a second source of reinforcement at work, the utility of correctly predicting the occurrence of the less frequent event, which subtracts from maximum gain responding (Brackbill, Kappy, & Starr, 1962). This also is consistent with the questionnaire responses; 77 of the experimental *Ss* reported that they knew the wisest procedure would be to bet consistently on the more likely alternative.

The data, thus, suggest a compromise between the expedience of betting 100% for the more likely alternative (because of the "chance" or "random" structure made patent by the experimental apparatus) and the challenge of betting 70% for the more likely alternative (i.e., trying out different patterns in an effort to "beat" the game).

Summary

The experiment investigated the effect of different levels of probability information on response frequencies in a random, two-choice situation with unequal event probabilities. This information was demonstrated to have differential effects during the early trials. In addition, it was found that *Ss* reached a significantly higher response level when the outcome of a trial appeared to depend upon the chance drawing of a marble from a box than when the marble box was absent. These findings were interpreted in terms of the perception of "randomness."

References

Anderson, N. H., & Grant, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.

Brackbill, Y., Kappy, M. S., & Starr, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63**, 32–35.

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Das, J. P. Mathematical solution in the acquisition of a verbal CR. *J. exp. Psychol.*, 1961, **61**, 376–378.

Edwards, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Flood, M. M. Environmental non-stationarity in a sequential decision-making experiment. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), *Decision processes.* New York: Wiley, 1954. Pp. 287–299.

Goodnow, J. J. Determinants of choice distributions in two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

Goodnow, J. J. A review of studies on probable events. *Aust. J. Psychol.*, 1958, **10**, 111–125.

Hake, H. W. The perception of frequency of occurrence and the development of "expectancy" in human experimental subjects. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955. Pp. 257–274.

Humphreys, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: A review and a critique. *Psychol. Bull.*, 1951, **41**, 291–297.

Koehler, J., Jr. Role of instructions in two-choice verbal conditioning with contingent partial reinforcement. *J. exp. Psychol.*, 1961, **62**, 122–125.

Lindman, H., & Edwards, W. Supplementary report: Unlearning the gambler's fallacy. *J. exp. Psychol.*, 1961, **62**, 630.

Neimark, E. D., & Shuford, E. H. Comparison of predictions and estimates in a probability learning situation. *J. exp. Psychol.*, 1959, **57**, 294–298.

Nicks, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

Rubinstein, I. Some factors in probability matching. *J. exp. Psychol.*, 1959, **57**, 413–416.

Siegel, S., & Goldstein, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

(Early publication received March 16, 1962)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 434-440

COLOR CODING AND VISUAL SEARCH[1]

SIDNEY L. SMITH

The MITRE Corporation, Bedford, Massachusetts

There is relatively little available data on the effectiveness of color as a coding dimension for information transmission in visual displays. Some studies have been published exploring the limitations of our ability to identify different hues on the basis of absolute discrimination. They suggest that under good viewing conditions 5 to 8 colors can be distinguished reliably (Conover, 1959; Conover & Kraft, 1958; Eriksen & Hake, 1955), or perhaps even as many as 9 to 12 under optimal conditions (Chapanis & Halsey, 1956; Halsey & Chapanis, 1951).

Whichever number we accept, this means that color as a coding dimension can be used to distinguish only a relatively small number of categories, as compared with shape coding (symbology) which carries the bulk of information in visual displays. However, within this limitation, there is evidence that color is better than shape in tasks which involve locating displayed data (Christner & Ray, 1961; Hitt, 1961).

The effectiveness of color coding for locating particular displayed items is related, by extension, to its use for providing visual separability *among* data classes. Results on this question have been reported by Green and Anderson (1956). They examined the degree to which color coding permitted visual separability of displayed two-digit numbers, as measured by decreases in average visual search time when the color of a "target" number was known beforehand (relevant) as compared with other occasions when it was not known (nonrelevant). They (Green & Anderson, 1956) summarized their conclusions as follows:

> When *Os* know the color of the target, the search time is approximately proportional to the number of symbols of the target's color. There is also a slight increment in search time due to the presence of the wrong-colored targets. When *Os* do not know the target's color, search time depends primarily on the total number of symbols on the display. However, search times are slightly longer for multicolored displays than for comparable single-colored displays (p. 24).

In an attempt to confirm these results, this present experimental study was conducted. This study accepts the basic premise of Green and Anderson that visual search time is a fundamental measure of the potential value of display color coding. It simply expanded their model to include a greater range of displayed densities, more displayed colors, both light and dark display backgrounds, and certain modified techniques of display presentation.

PROCEDURE

Twelve *Ss* participated in this study, 11 men and 1 woman. Preliminary testing using the American Optical Company H-R-R pseudoisochromatic plates confirmed that all *Ss* had normal color vision. In the course of the study, each *S* made a total of 300 visual searches, using a variety of different displays. This required several experimental sessions for each *S*. Individual *Ss* worked for no longer than 1 hr. at a time, and for no more than 2 hr. per day.

The displays consisted of varying arrays of three-digit numbers. These numbers were

[1] The research reported in this article was supported by the Department of the Air Force under Contract AF-33(600)39852. A more detailed account of this research was published as a MITRE Technical Series Report, MTS-7, "Display Color Coding for a Visual Search Task," June 1962.

randomly placed in a square field, which can be imagined as comprising 13 columns and 27 rows, for a total of 351 possible positions. The numbers themselves were chosen randomly from the 1000 possibilities (000 through 999) with certain restrictions: the numbers on each particular display were all different; they were unique in terms of their first two digits; their third digits represented an equal sample of each of the 10 possibilities (0 through 9). In the case of multicolored displays, the particular color of each three-digit number was also chosen at random, with the restriction that all colors were equally represented.

The displays were made up as 2 × 2 in. color slides, and presented to *Ss* by rear projection on the screen of an experimental console. The dimensions of the digits as projected on this screen were ½ in. high × ¼ in. wide. Viewing distance was about 18 in. The overall display field as projected was 12 in. square.

The colors used on the various displays were red, green, blue, orange, and either white (on slides with a black background) or black (on a white background). Visual matches of the colors as projected with standard colors resulted in agreement by 2 *Os* on the descriptive specifications shown in Table 1.

Because rear projection was used, a moderately high ambient illumination was maintained, by diffuse overhead lighting from dimmed fluorescents: over .5 ft-c as measured at the experimental console.

The experimental routine for *S* began when *E* signaled to him the first two digits of the target number, the number he was to find on the display. These two digits were displayed on an auxiliary panel. The *S* also saw on this panel either a colored indicator showing what color the target number would be, or else a statement "color unknown." The *E* then exposed the slide on the screen in front of *S* and started a clock. The *S* searched for the target number, indicated when he had found it by pushing one of 10 numbered buttons, corresponding to the third digit of the target number. If this response was correct, the clock stopped, a chime sounded, and the screen went blank. The *E* recorded the time, and then set up the auxiliary display panel to begin the next trial. In the very rare case when *S* pushed the wrong response button, a loud buzzer sounded, and the trial was run again later in the experimental series.

TABLE 1

DISPLAY COLORS USED (MUNSELL NOTATION)

Display Color	White Background	Black Background
Red	2.5 R 5/10	5 R 5/12
Green	7.5 GY 8/8	5 GY 7/8
Blue	2.5 PB 6/8	2.5 PB 6/8
Orange	2.5 YR 7/10	2.5 YR 7/10
Black/White	7.5 P 3/4	N 9/0

TABLE 2

DISPLAYS USED OF EACH TYPE

Number of Colors Displayed	Number of Displayed Items				
	20	40	60	80	100
1	5	5	5	5	5
2		10			
3			10		
4				5	
5	5	5	5	5	5

The displays were presented in a series of 150 slides, half with white backgrounds and an identical group with black backgrounds. Each group of 75 slides contained both single- and multicolored displays with display densities varying from 20 to 100 three-digit numbers (which will be called "items" in the next few paragraphs to avoid confusion). The types of displays used are summarized in Table 2. The entries in the matrix represent the number of different displays used.

There are really three sets of displays represented in the matrix. The upper row consists of single-colored displays of increasing display density, with 5 displays (one of each color used) at each density level. The lower row consists of a comparison set of displays, each with items of all five colors on it. It was arbitrarily decided to use 5 displays at each density level, to match the first set. The diagonal set of the matrix represents those displays where an increasing number of displayed colors is associated with increasing display density, with the advantage that a constant number (20) of items of any particular color are displayed amid varying numbers of other-colored items: 10 two-colored displays were used, representing all the possible pairs of the five colors available; 10 three-colored displays represented all the possible triple combinations; and 5 four-colored displays represented the cases where each of the five available colors had been omitted.

The series of 150 displays described above

were presented in a random order, which in turn was randomized (in five blocks) differently for different *Ss*. Each *S* went through this series twice in the course of the experiment, for a total of 300 searches. The first time through, for half of the displays (randomly chosen) he was told in advance the color of the target number, the number-to-be-searched-for. For the other half of the displays, the color of the target number was unknown to him. During his second run through the series, these conditions were reversed for each particular display—where the color of the target had been known before, it was now unknown, and vice versa.

In summary, visual search time data were obtained from 12 *Ss*, each viewing a series of 300 displays, which varied in display density, in number of colors used, in the particular color of the target, with either a white or black background, under conditions where *S* either knew the color of the target number in advance or did not.

Results and Discussion

Because of the relative complexity of the experimental design, no single statistical treatment would suffice to answer all the questions we might wish to raise. In fact, five separate analyses of variance were made, each using some portion of the data, in order to examine the effect of different combinations of the experimental variables.

The most extensive analysis of variance, and one basic to the subsequent data analysis, was carried out using individual search times obtained under all conditions involving either single-colored or five-colored displays. These data represent a factorial design of 12 *Ss* by 2 conditions of prior knowledge (target color either known or unknown) by 2 types of display background by 5 degrees of display density by 2 types of display (either single-colored or five-colored) by 5 possible colors used for the target number. Taken together they comprise 2400 measures.

The data initially available for this analysis were search times expressed in .01 sec. Because of the inherently high correlation between mean and variance in search time data, a log transform was used prior to the variance analysis computations. The computations themselves were carried out by an electronic computer, and sums of squares were obtained for the six experimental variables and their 57 various interaction terms, including the six-way residual. Following a procedure suggested by Edwards (1950) for the treatment of repeated measures obtained from the same *Ss*, all sums of squares representing interactions of *Ss* with other experimental variables were pooled to form one residual representing the *Ss* × Conditions interaction. The mean square of this interaction was used as the error term to test the significance of all others. (As it happens, if *Ss had* been treated as a legitimate variable in the analysis, none of the interactions involving *Ss* would have proved statistically significant.) Because of the large number of *F* ratio comparisons made in this analysis, only significance levels of at least .001 were accepted as persuasive evidence of statistical reliability. The summarized results of this analysis are presented in Table 3.

We may note, first, that neither the particular target color used, nor the display background, nor any interaction term including these variables, had any statistically significant effect. We might have expected, for example, that some colors on a light background would be less legible than on a dark background, and hence more difficult to scan quickly, since it is clear that visual contrast is an important variable in legibility. However, the present data suggest that we need not expect any measurable effect related to different contrast ratios

TABLE 3

ANALYSIS OF VARIANCE OF TRANSFORMED SEARCH TIMES FOR ALL EXPERIMENTAL CONDITIONS INVOLVING ONE- AND FIVE-COLORED DISPLAYS

Source of Variance	*df*	*MS*	*F*
Ss	11	0.708	—
Conditions	199	—	—
Knowledge of target color (K)	1	21.787	229.34*
Display background (B)	1	0.301	3.17
Display density (D)	4	16.791	176.75*
Number of colors (N)	1	14.956	157.43*
Color of target (C)	4	0.063	0.66
K × B	1	0.015	0.16
K × D	4	0.619	6.52*
K × N	1	15.549	163.67*
K × C	4	0.106	1.12
B × D	4	0.071	0.75
B × N	1	0.000	0.00
B × C	4	0.095	0.98
D × N	4	0.129	1.36
D × C	16	0.129	1.36
N × C	4	0.292	3.06
Higher-order interactions among conditions	145	0.096	—•
Ss × Conditions	2189	0.095	—

• None was significant when tested individually.
* $P < .001$.

that are all well above some sort of legibility threshold.

Next, consider the marked effect on search time of increasing display density, confirmed by the statistical analysis and illustrated in varying degrees by Fig. 1–3. It is the nature of these data that there is greater inherent variability of the search time measure in those situations where the average search takes longer. Taking this into account, it seems that the average search time data closely approximate a direct linear relation with display density. This was also the case in the Green and Anderson study and has been noted before by other investigators (e.g., Green, McGill, & Jenkins, 1953).

If the "curves" in Fig. 1 are extrapolated backward to a display density of zero, they would intersect the ordinate axis at a value of about 1 sec., which presumably represents the simple button-pushing reaction time. This is confirmed by the observation that only one or two of the individual search time measures obtained in this study were less than 1 sec., and those were smaller by only a slight margin. The same observation was made by Green and Anderson (1956) under similar circumstances. It should be pointed out that their reported search times were in general somewhat shorter than those in the present study. In part, this is because they reported geometric rather than arithmetic means. And, in part, this difference is probably attributable to the fact that their display field ($12 \times 16\frac{3}{4}$ in., viewed from a distance of 10 ft.) subtended a smaller visual angle than was the case in this present study. Hence, fewer eye movements were required to scan their displays.

The significant interaction between knowledge of target color, and number of colors displayed, indicates the potential value of color coding when the color is relevant to the search task. This is illustrated by the sizable difference between the curves in Fig. 1. If the color coding permitted absolute visual separability between classes of displayed items, then the average search times for these five-

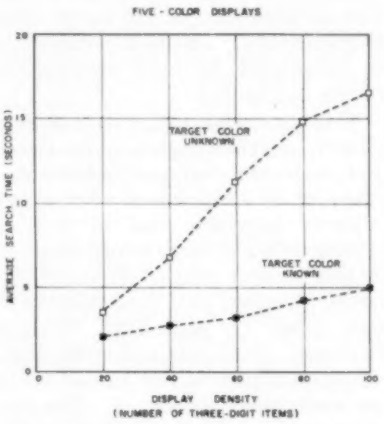

FIG. 1. Search time as a function of display density with knowledge of target color as a parameter.

color displays would be only one-fifth as great when the target color was known as when it was unknown. In actuality, the difference was not quite so great as this, but considerable nonetheless.

Comparable curves for single-colored displays showed no such difference, i.e., knowing the target color beforehand did not speed the subsequent visual search process. Indeed, there is no logical reason to expect that it would. However, to check this, a separate variance analysis was conducted, similar to that already described, but using only the search times for single-colored displays. The only significant variable in this treatment turned out to be display density. Prior knowledge of the target color made no difference.

A comparable variance analysis was run comparing one- and five-colored displays, but using only the search time data obtained when *S*s had no prior knowledge of target color. Again, the only significant variable proved to be display density. In particular, number of colors displayed made no difference to *S*s when they had no knowledge of the target color to guide them. This lack of difference is illustrated by the overlapping curves in Fig. 2.

This last finding would seem to bear directly on the conclusion by Green and Anderson that multicolored displays retard visual search when target color is unknown, and in fact, is evidence that no such effect occurred in this present study. It should be noted, however, that their conclusion was based on somewhat different evidence. Data obtained under equivalent conditions in this present study are summarized in Fig. 3. The displays considered here are those which always contained 20 items of the same color as the target number. For the

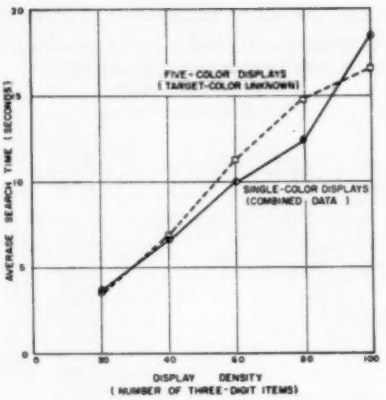

FIG. 2. Search time as a function of display density under conditions of nonrelevant color coding and of no color coding (single-colored displays).

displays with just one color, this is all they contained. For the two-colored displays, there were also 20 items of the second color present. For the three-colored displays, there were 20 items displayed for *each* of the two other (nontarget) colors—and so on, for the four- and five-colored displays.

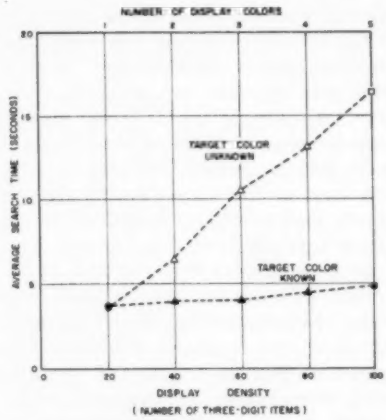

FIG. 3. Search time as a function of display density when target color is known and unknown and when there are 20 items of target color in all cases.

This comparison in Fig. 3, showing the considerably reduced search times when the target color was known to *Ss*, is essentially the comparison made by Green and Anderson in their report. It is true that the number of displayed colors is here confounded with differences in display density. But in the sense that each display contains just 20 target numbers amid varying amounts of "clutter," the comparison it permits is interesting.

To permit a direct check of the Green and Anderson (1956) results, a further analysis of variance was run, comparing search times for single-colored displays, of 40, 60, 80, and 100 items, with those for two-, three-, four-, and five-colored displays, of corresponding respective densities, under conditions of no prior knowledge of target color. Since neither display background nor target color had proved to be significant variables in preceding analyses, these were eliminated in this analysis by summing individual search times across these variables. Thus the data used consisted of the sum of 10 search times (5 target colors by 2 display backgrounds) under each combination of display density and number of colors. Actually, this is not entirely true, for in the case of the two-colored, 40-item displays, and the three-colored, 60-item displays, there were 20 search times available (10 target color/other color combinations by 2 display backgrounds). In these cases, half the available data were chosen randomly to use in the variance analysis. This total data subset, then, represents a factorial design of 12 *Ss* by 4 degrees of display density by 2 types of displays—single-colored versus multicolored—for a total of 96 measures.

The results of this analysis are summarized in Table 4. The only variable with a statistically significant effect proves again to be display density. There is no reliable difference that can be attributed to number of displayed colors. If such a difference had been confirmed, it would have been in favor of the multicolored displays, which were searched in an average time of 5.9 sec., as compared with 6.5 sec. for the single-colored displays. Certainly we cannot conclude that multicolored displays were distracting in this present study, even though the target color was unknown.

TABLE 4

ANALYSIS OF VARIANCE OF SUMMED SEARCH TIMES FOR CONDITIONS INVOLVING SINGLE- AND MULTICOLORED DISPLAYS WITH NO PRIOR KNOWLEDGE OF TARGET COLOR

Source of Variance	*df*	*MS*	*F*
Ss	11	9,351	—
Conditions	7	—	—
Number of colors (N)	1	3,449	0.71
Display density (D)	3	68,973	14.11*
N × D	3	4,397	0.90
Ss × Conditions	77	4,888	—

* $P < .001$.

Why did Green and Anderson obtain a different result? The most probable explanation lies in their choice of projection technique: luminous colored numbers displayed in a dark environment would provide a stimulus situation very conducive to perception of a depth illusion based on differences in either brightness or wave length. If their displayed numbers of different colors appeared to be in slightly different frontal planes, this might have led their *Ss* to scan a multicolored display several times, a process that on the average would be somewhat slower than one systematic search of a single-colored display. In this present study, with rear projection and higher ambient illumination, the displayed symbols clearly lay on a single surface, and no depth effects were apparent.

The slight but regular increase in search time, when the target color was known, as more and more numbers of other colors were added to the display (the lower curve in Fig. 3), also proved to be statistically reliable. The data involved in this demonstration consist of the search times for each individual *S*, summed over target color and display background as before, for displays of one, two, three, four, and five colors, with display densities of 20, 40, 60, 80, and 100 items, respectively, under conditions where the target color was known in advance. As in the previous analysis, only half of the available data for the two-colored, 40-item displays and the three-colored, 60-item displays was used. This represents a factorial combination of 12 *S*s by 5 display densities (each representing a different number of displayed colors), for a total of 60 measures. Analysis of variance confirmed a reliable difference among the 5 display types at $P < .01$ ($F = 4.37$; $df = 4/44$). This confirms the conclusion of Green and Anderson, based on a similar analysis, that wrong-colored items can be almost completely ignored in visual search—almost, but not quite.

Summary

Twelve *S*s each viewed a series of 300 displays, which varied in display density, in number of colors used, in the particular color of the target, with either a white or black background, under conditions where *S* either knew the color of the target in advance, or did not.

Neither the particular color of the target nor the display background had any significant effect on search time. Search time increased regularly with increasing display density. For multicolored displays, when the color of the target was known in advance, search times were considerably shorter than when the target color was unknown. When the color of the target was unknown, search times were not significantly different than those for single-colored displays.

References

Chapanis, A., & Halsey, R. M. Absolute judgments of spectrum colors. *J. Psychol.*, 1956, **42**, 99–103.

Christner, C. A., & Ray, H. W. An evaluation of the effect of selected combinations of target and background coding on map-reading performance: Experiment V. *Hum. Factors*, 1961, **3**, 131–146.

Conover, D. W. The amount of information in the absolute judgment of Munsell hues. *USAF WADC tech. Note*, 1959, No. 58–262.

Conover, D. W., & Kraft, C. L. The use of color in coding displays. *USAF WADC tech. Rep.*, 1958, No. 55–471.

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

Eriksen, C. W., & Hake, H. W. Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, 1955, **50**, 153–160.

Green, B. F., & Anderson, L. K. Color coding in a visual search task. *J. exp. Psychol.*, 1956, **51**, 19–24.

Green, B. F., McGill, W. J., & Jenkins, H. M. The time required to search for numbers on large visual displays. Technical Report No. 36, 1953, Massachusetts Institute of Technology, Lincoln Laboratory.

Halsey, R. M., & Chapanis, A. On the number of absolutely identifiable hues. *J. Opt. Soc. Amer.*, 1951, **41**, 1057–1058.

Hitt, W. D. An evaluation of five different abstract coding methods: Experiment IV. *Hum. Factors*, 1961, **3**, 120–130.

(Early publication received April 24, 1962)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 441–450

RESISTANCE TO EXTINCTION WHEN PARTIAL REINFORCEMENT IS FOLLOWED BY REGULAR REINFORCEMENT

HERBERT M. JENKINS

Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey

The discrimination hypothesis (Humphreys, 1940; Mowrer & Jones, 1945) ascribes the partial reinforcement effect (PRE) to the relative ease of discriminating an abrupt transition to extinction following regular reinforcement as compared to the more gradual transition following partial reinforcement.

If the recent conditions of reinforcement provide the basis of the discrimination, then the PRE should be decreased by the interpolation of regular reinforcement prior to extinction since this would make the transition between recent training and extinction abrupt. This was examined for the case of the free operant response by Keller (1940), by Likely (1958), and by Quatermain and Vaughan (1961). Although in every case the results were that the interpolation of regular reinforcement failed to reduce resistance to extinction, no firm conclusions can be drawn since none of these experiments demonstrated a clear PRE for partial versus regular reinforcement taken separately. Theios (1962) remedied this deficiency in a runway experiment. The PRE was not significantly altered by the interpolation of 25 regular reinforcements, and it was still clearly present, although significantly reduced, after 70 regular reinforcements.

The present experiments use the pigeon's key peck response with a discrete trial procedure. They were carried out independently of Theios' work and afford some of the same comparisons. In Exp. I, the location of a period of partial reinforcement within a longer regime of training under regular reinforcement was varied so that the effect of the amount of regular reinforcement which occurred between partial reinforcement and extinction could be observed where the total amount of training was the same. A baseline was provided by a control group which received all regular reinforcement. The results suggested that the interpolation of regular reinforcement prior to extinction may *increase* resistance to extinction over the level obtained when extinction occurs directly following partial reinforcement. Experiment II was designed to substantiate this finding, to further define the conditions under which it occurs, and to remove certain ambiguities in the interpretation of Exp. I.

METHOD

Subjects.—The *S*s were 5–6 yr. old, male White Carneaux pigeons without prior experimental history. They were maintained by restricted feeding at 80% of their free-feeding body weight.

Apparatus.—An automatic key pecking apparatus of Skinner design was used. Details of the apparatus have been reported previously (Jenkins, 1961). Reinforcement was a 4-sec. period of access to a tray of mixed grain signaled by lighting the opening to the tray.

Trials.—A trial was begun by lighting the translucent plastic response key. The trial was terminated (key light turned off) by a single response, or by external control at the end of 5 sec., whichever occurred first. On reinforced trials, the tray operation followed the response without delay. During the

interval between trials the key light was off, but *S*'s compartment remained illuminated. The time between onsets of successive trials was equally often 15, 30, or 45 sec. in an irregular order. A response made between trials prevented the onset of the next trial for 60 sec. A second response occasionally occurred closely following the first response to a trial. On trials in which the first response initiated an operation of the grain tray, the second response released the tray and thus prevented reinforcement. The second response always placed the 60-sec. delay into effect.

Two measures were recorded on each trial: (*a*) the occurrence or nonoccurrence of a response within the 5-sec. trial period, and (*b*) the latency of response in tenths of a second. No latency was entered for trials on which *S* failed to respond. Intertrial responses (ITRs) were also recorded.

Shaping of response.—The *S*s were trained to strike the lighted key by reinforcing successive approximations. When several automatically reinforced key responses had occurred, intertrial intervals were introduced. A discrimination between the lighted and unlighted key was formed immediately so that responses were confined almost entirely to the trials from the beginning of training. Two sessions were, with a few exceptions, sufficient to prepare *S*s for the automatically programed training.

Sessions.—The experiments involved preliminary training, training, and extinction. Preliminary training consisted of four sessions (Exp. I) or five sessions (Exp. II) of regular reinforcement. In each session, 40 reinforced trials were programed. The period of access to the grain tray was reduced from 5 to 4 sec. in the course of these sessions.

Training sessions were under regular or partial reinforcement. In sessions of regular reinforcement 40 reinforced trials were programed as in preliminary training. Sessions of partial reinforcement consisted of 40 reinforced trials plus a number of nonreinforced trials programed in a random sequence subject to the constraint that each half of the sequence contained half of the reinforced trials (20) and half of the nonreinforced trials.

Extinction sessions were run on consecutive days and each consisted of 40 nonreinforced trials. By prior decision, all comparisons among groups were based on the first 10 extinction sessions. Additional sessions were run in the case of most groups in order to more nearly complete extinction.

Treatment groups, Exp. I.—The four groups of Exp. I each received 13 sessions of training beyond preliminary training. Group 13R received 13 sessions of regular reinforcement. The remaining three groups received partial reinforcement for a block of 3 consecutive sessions located in different positions within the sequence of training sessions. These groups are designated 3P-10R (3 sessions of partial reinforcement followed by 10 sessions of regular reinforcement), 9R-3P-1R, and 10R-3P. Forty nonreinforced trials were programed in the first session of partial reinforcement, 60 in the second, and 80 in the third.

Five *S*s were assigned to each group so as to match the groups as well as possible on the basis of the means and variance of the latency of response during preliminary training.

Treatment groups, Exp. II.—Three groups received different numbers of training sessions under partial reinforcement: Groups 8P ($N = 9$), 20P ($N = 7$), and 32P ($N = 7$). Two groups received different amounts of training under regular reinforcement: Groups 8R ($N = 6$) and 20R ($N = 6$). Finally, two groups received different amounts of training under partial reinforcement prior to 12 sessions of regular reinforcement: Groups 8P-12R ($N = 7$) and 20P-12R ($N = 6$). The reasons for selecting these conditions will become clearer after the results of Exp. I are reported, but certain features may be noted here. Partial reinforcement was begun at the same point in training (directly following preliminary training), rather than at different points as in Exp. I. The amount of training was extended in order to obtain a clear separation in resistance to extinction between groups given partial reinforcement as against those given regular reinforcement. The design also allows comparison of the effect on extinction of adding, after different amounts of training under partial reinforcement, either more partial reinforcement, or an equal number of sessions of regular reinforcement.

The groups in Exp. II were run in the following order: Groups 8P and 8R concurrently, then Groups 20P, 20R, and 8P-12R concurrently, then Groups 32P and 20P-12R concurrently. Groups run at the same time were matched on the basis of performance in preliminary training as in Exp. I.

In each of the first two sessions of partial reinforcement, 40 nonreinforced trials were programed. Thereafter, each session of partial reinforcement contained 80 nonreinforced trials.

Additional sessions of extinction.—All groups in Exp. I and Groups 20R, 20P, and 8P-12R of Exp. II received 2 additional sessions of extinction. Groups 20P-12R and

32P received 4 additional sessions of extinction. In Group 32P 1 *S* was given 5 and another was given 11 sessions beyond Session 10. With the additional sessions of extinction, there was at least 1 session for every *S* in which 10 or fewer responses were made. The effect of including responses made in sessions beyond Session 10 on the relative standing of the groups as determined by the first 10 sessions was examined and is reported below.

Results

Responses in training.—The proportion of trials responded to was maintained at very high levels throughout all phases of training. The mean proportion based on both experiments was .976 for preliminary training and .985 for training. There were no systematic differences in this measure among groups or between the periods of partial reinforcement and regular reinforcement within groups.

Shape of extinction curves.—The discrete trial method of the present experiments yields extinction curves which typically show a sharp drop from a high to a low level of responding. Individual extinction curves are shown for *S*s in Groups 20R and 20P in Fig. 1. Although the average number of responses in extinction was about three times greater in Group 20P than in Group 20R, it appears from Fig. 1 that the curves differ principally in the points at which failures to respond begin, rather than in the abruptness of the decline thereafter.

In order to examine the abruptness of extinction more systematically, two criterion sessions designated A and B were located for each *S*. Session A was the last one in which 30 or more responses were made and Session B was the first one in which 10 or fewer were made. Figure 2 plots for each group in Exp. II the mean locations (session numbers) of the

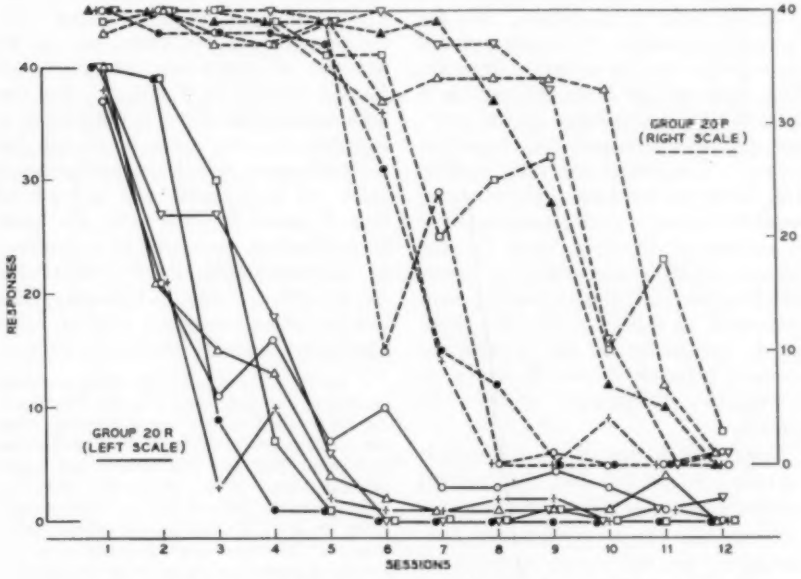

FIG. 1. Responses in extinction for individual *S*s of Groups 20P and 20R.

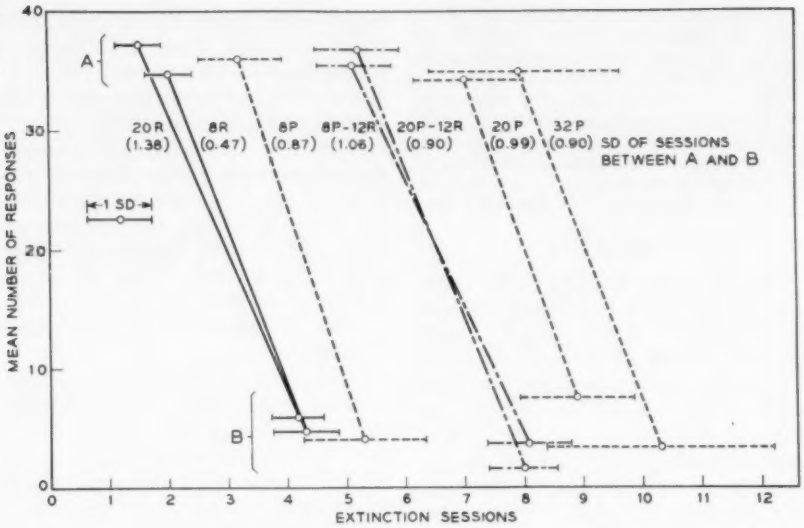

FIG. 2. The average slopes of extinction curves for the groups of Exp. II. (The criterion sessions A and B are explained in the text. The bars represent one *SD* of the location of Sessions A and B, or one-half *SD* to either side of the mean. The numbers in parentheses are one *SD* of the number of sessions between Sessions A and B.)

criterion sessions A and B. On the ordinate is plotted the mean number of responses in the criterion sessions. The slope of the lines connecting A with B thus represents, on the average, the abruptness of the extinction curves. The slopes are very similar. The *SD* of the location of the criterion sessions increases as the mean number of extinction sessions prior to the "break point" increases, a trend which is correlated with partial reinforcement in training. On the other hand, variability in the number of sessions between A and B shows no systematic differences among the groups.

Since the shapes of the extinction curves are similar for the several groups, the major effects of the training variables can be represented conveniently by the mean number of responses in extinction.

Responses in extinction, Exp. I.— Summary data for responses in 10 sessions of extinction are given for Exp. I and II in Table 1.[1] On the assumption that PRE is reduced as a function of the amount of regular reinforcement which is interpolated prior to extinction, the groups of Exp. I would be ordered on the basis of number of responses in extinction as follows: 10R-3P > 9R-3P-1R > 3P-10R ≥ 13R. However, the groups which received partial reinforcement actually show the reverse

[1] The inclusion of responses which occurred in sessions of extinction beyond the tenth session resulted in only minor increases over the totals given in Table 2 except in the case of Groups 20P and 32P, where the totals increase about 10%. Statistical tests involving Groups 20P and 32P were recomputed using the total of responses for all extinction sessions. The significance levels in each case remained unchanged from those obtained on the basis of the first 10 sessions of extinction.

ordering: i.e., 10R-3P < 9R-3P-1R < 3P-10R. An analysis of variance for these groups yielded an $F = 3.40$ with 2/12 *df*, and $P < .10$. The mean difference between the extreme groups was 78.4 responses ($t = 2.60$; $df = 8$, $P < .05$). Group 13R, which received no partial reinforcement had the lowest mean number of responses in extinction. However, meaningful statistical comparisons of this group with the others are difficult to make since 1 *S* in Group 13R resumed responding late in extinction and thus made a total number of responses exceeded by only 2 other *S*s in the entire experiment. The remaining 4 *S*s in Group 13R had the four lowest totals in the experiment.

The following hypothesis was formulated from the outcome of Exp. I and tested in Exp. II: The interpolation of regular reinforcement between partial reinforcement and extinction does not reduce the PRE, but in fact increases resistance to extinction over that which obtains when extinction follows partial reinforcement directly. Experiment I falls short of establishing this point on two counts. First, a clear PRE was not obtained when extinction followed partial reinforcement directly (i.e., Groups 10R-3P and 13R were not well separated). In order to obtain a clear PRE against which to evaluate the effect of interpolated regular reinforcement, the amount of training under regular or partial reinforcement was extended in Exp. II. Second, although the hypothesis ascribes the greater resistance to extinction in Group 3P-10R compared to Group 10R-3P to the regular reinforcement which followed partial reinforcement, the difference between these groups could also be ascribed to differences in the amount of regular reinforcement which preceded the partial reinforcement. In Exp. II, partial reinforcement was introduced at the same point in training in order to avoid this ambiguity.

Responses in extinction, Exp. II.—A comparison of Groups 8P, 20P, and 32P of Exp. II shows that resistance to extinction was a function of the amount of training under partial reinforcement ($F = 12.18$; $df = 2/12$, $P < .001$). The extension of training from 8 to 20 sessions resulted in a significant increase in resistance to extinction ($t = 4.16$; $df = 14$, $P < .001$), but no further increase resulted from the further extension of training from 20 to 32 sessions ($t < 1$).

TABLE 1

MEANS AND *SD*s OF RESPONSES IN 10 SESSIONS OF EXTINCTION: EXP. I AND II

Type of Training	Experiment I			Experiment II		
	Group	Mean	*SD*	Group	Mean	*SD*
Regular	13R	100.0	71.6	8R	114.5	33.2
				20R	103.7	15.1
Partial	10R-3P	118.4	48.2	8P	164.3	63.9
				20P	294.9	54.9
				32P	298.7	54.9
Partial then Regular	9R-3P-1R	162.2	41.2	8P-12R	247.4	34.7
	3P-10R	196.8	37.8	20P-12R	250.2	48.3

A comparison of Groups 8R and 20R shows, on the other hand, that when training was under all regular reinforcement, resistance to extinction was unchanged by the extension of training from 8 to 20 sessions ($t < 1$).

A clear PRE emerged only after extended training. Groups 8R and 8P did not differ significantly ($t = 1.64$; $df = 13$, $P < .20$) while Groups 20R and 20P were clearly different ($t = 7.60$; $df = 11$, $P < .001$).

Consider next Group 8P-12R which was switched from partial to regular reinforcement prior to extinction. Resistance to extinction for this group was significantly greater than for Group 8P ($t = 2.91$; $df = 14$, $P < .02$) showing that the addition of regular reinforcement increased resistance to extinction. The increase obtained from adding 12 sessions of regular reinforcement to 8 sessions of partial reinforcement was less, but not significantly less, than the increase resulting from the addition of 12 sessions of partial reinforcement, i.e., the means for the number of responses in extinction in Groups 8R-12P and 20P could not be reliably distinguished ($t = 1.79$; $df = 12$, $P \cong .10$).

When 20 sessions of partial reinforcement were given, the addition of 12 sessions of regular reinforcement (Group 20P-12R) produced no further increase in resistance to extinction. The mean number of responses to extinction was in fact less, although not significantly less, in Group 20P-12R than in either Groups 20P or 32P (20P-12R vs. 20P: $t = 1.42$, $df = 11$, $P < .20$; 20P-12R vs. 32P: $t = 1.55$, $df = 11$, $P < .20$).

The results of Exp. I and II are brought together in Fig. 3 which plots the mean number of responses in extinction as a function of the amount of training under regular reinforcement, the amount of training under partial reinforcement, and for the mixed conditions in which partial reinforcement is followed by regular reinforcement. Experiment I has been coordinated to this plot by ignoring differences in the amount of regular reinforcement which preceded the introduction of partial reinforcement. When treated in this way, the outcome of Exp. I falls into line quite well with that of Exp. II, so that it seems reasonable to believe that variations in the amount of training under regular reinforcement preceding the introduction of partial reinforcement made little difference.

The independent variables in Fig. 3 are the number of training sessions of regular or partial reinforcement *beyond* the sessions of preliminary training. Thus, the dotted lines extrapolated to zero sessions of training represent a guess that preliminary training alone would produce a resistance to extinction equal to that obtained after eight additional sessions of regular reinforcement.

Relation of latency in training to responses in extinction.—The amount and type of training had only minor effects on the mean latency of response for groups. There were in fact no significant differences in the mean latencies of response taken over the last three sessions of training in either experiment. The overall mean latency for these sessions was 1.4 sec. However, it was possible to detect in Exp. II a small effect of partial reinforcement when compared to regular reinforcement for the first eight training sessions. The mean latency for partially reinforced *S*s ($N = 36$) was .2 sec. longer than for regularly reinforced *S*s ($N = 12$). A test of this difference using a covariance analysis with latency during preliminary training as the predictor yielded $t = 2.34$; $df = 45$, $P < .05$. The switch from

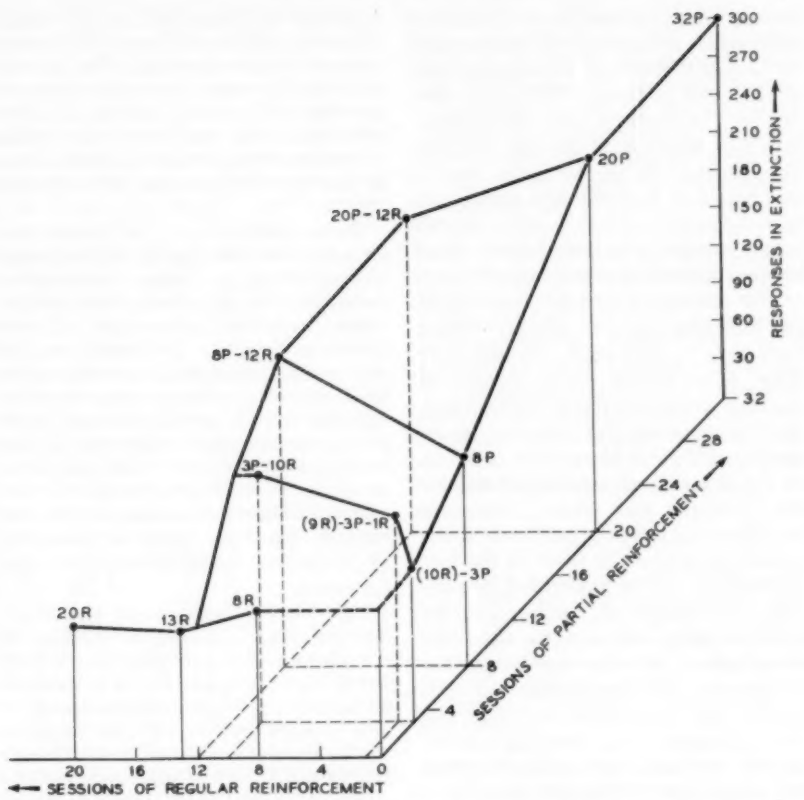

FIG. 3. Mean number of responses in extinction for all groups (Exp. I and II) as function of the amount and type of training beyond preliminary training. (Sessions of regular training preceding partial training in Exp. I have been disregarded and are shown in parentheses [see text].)

partial to regular reinforcement caused no significant change in latency in either experiment.

From the overall results it is apparent that the effects of the amount and type of training on resistance to extinction were not paralleled by effects on group mean latencies of response in training.

The correlation between latencies for individuals within groups and responses to extinction is more interesting since it provides a clue as to how the addition of training under regular reinforcement after a limited amount of training under partial reinforcement operates to increase resistance to extinction. A substantial negative correlation (longer latencies associated with fewer responses in extinction) was found in the case of two groups: 10R-3P (Exp. I) and 8P (Exp. II). These were the only groups which went directly to extinction after a training period under partial reinforcement that was

not sufficiently extensive to produce a maximum resistance to extinction. In Group 10R-3P the correlation was $-.79$, and in Group 8P it was also $-.79$. In the first case the correlation was based on 5 *S*s and was not significant while in the latter case it was based on 9 *S*s and was significant ($t = 3.35$; $df = 7$, $P < .02$). In all other groups, the correlations were low and did not approach significance.

The pattern of correlations reflects the following state of affairs. When extinction was begun immediately after a relatively short period of training under partial reinforcement, slow responders extinguished more rapidly than fast responders. When, on the other hand, a period of regular reinforcement was added, resistance to extinction for slow responders was at the same average level as for fast responders. It is important to note that the effect of adding regular reinforcement was not to alter the mean latency of response, but rather to remove the correlation between latency and resistance to extinction. The extension of training under partial reinforcement also removed the correlation in the same way.

Discussion

The abruptness of the transition from training to extinction is obviously not the critical functional difference between regular and partial reinforcement. When extensive training is first given under partial reinforcement, its effect persists through a period of training under regular reinforcement involving 480 reinforcements.

Although 480 reinforcements is a large number, it is still a small fraction of the 2,320 training trials under partial reinforcement which were used in Group 20P in order to obtain a clear PRE. It therefore seems possible that the insignificantly lower resistance to extinction for Group 20P-12R in comparison with that of Group 20P or 32P would become a reliable difference with more extensive regular training. The point is important in connection with one interpretation of how overtraining in a discrimination task facilitates the learning of a reversed discrimination (Birch, Ison, & Sperling, 1960; Capaldi & Stevenson, 1957).

Interpolating a period of regular reinforcement between partial reinforcement and extinction can increase resistance to extinction over the level which obtains when extinction follows partial reinforcement directly. The results suggest the rule that added training under regular reinforcement increases resistance to extinction only if added training under partial reinforcement would also increase resistance to extinction. It appears that *S*s with relatively long latencies of response during the first segment of partial training gain the most in resistance to extinction from subsequent reinforcement.

This unanticipated result requires a reformulation. Instead of thinking of the partial training as producing a PRE which might or might not be attenuated by subsequent regular reinforcement, it now appears that the PRE can be *generated* by the training conditions in Groups 3P-10R and 8P-12R. The present interpretation of the PRE in these groups is that it arose from an interaction of the effects of the nonreinforced trials during the partial training with the subsequent reinforcement during regular training. It is clear that the effect depends upon an interaction since without a prior exposure to nonreinforced trials, resistance to extinction was unchanged by the extension of training under regular reinforcement. Amsel's (1958) account of the PRE, which holds that the combination of conditioned frustration due to nonreinforcement and subsequent reinforcement develops a tolerance for nonreinforcement, provides one conjecture as to how this interaction is mediated.[2]

[2] Results on the occurrence of ITRs suggest that frustration was associated with extinction. During training, the overall mean

Other results (Weinstock, 1954, 1958) indicate that some form of interaction between nonreinforcement and reinforcement can occur across long time intervals (at least 24 hr. separated the partial from the regular training in the present experiments), and even when the reinforcements occur consecutively in a block (Jensen & Cotton, 1960; Lauer & Carterette, 1957) as they did in the regular training of the present experiments.

The present results agree with those of Keller (1940), Likely (1958), Quatermain and Vaughan (1961), and more specifically with Theios (1962), in demonstrating the persistence of the PRE through a period of regularly reinforced training. Further, Theios' finding of a small but significant reduction in the PRE for the group which received more extensive regular reinforcement (Group P-70) is paralleled by the present finding of a lessened resistance to extinction in Group 20P-12R.

The fact of increased resistance to extinction for the partial-regular training sequence is a departure, although indications of it can be found in previous results. In particular, Theios obtained a small increase in resistance to extinction (not significant) when 25 regular reinforcements were added to previous partial training. He considered the possibility that it resulted from the regular reinforcement, but rejected it on the grounds that no increase occurred when regular reinforcement was added to previous regular training. However, the present results show that when the effect is obtained, it arises from an *interaction* of partial and regular reinforcement; a possibility which is overlooked in Theios' argument.

Two implications of the present results for theory may be noted. First, the increased resistance to extinction produced by the partial-regular sequence poses a problem for the dissonance theory of the PRE (Lawrence & Festinger, 1962) since that theory holds that the processes producing the increased resistance to extinction occur primarily on the nonreinforced trial itself. Second, the remarkable degree of independence between the slope of the extinction curves and the total number of responses in extinction indicates that something like a threshold model of extinction is preferable, at least for conditioning procedures of the type used here, to a decremental model in which an initial probability of response is reduced on each trial by an amount proportional to the number of responses remaining. Models of the latter type lead to exponential decays in which the slope of the curve is correlated with the total number of responses in the curve.

frequency of ITRs was .5 per 100 trials. In extinction it increased in every group yielding an overall mean of 2.2 per 100 trials. The mean frequency of ITRs showed a peak value in the vicinity of the decline of responses in extinction.

Summary

A food reinforced, key peck response in the pigeon was used with a discrete trial procedure to study resistance to extinction (number of responses in 10 40-trial sessions) following different amounts of training of three types: partial reinforcement, regular reinforcement, and partial followed by regular reinforcement.

Resistance to extinction showed no systematic change as a function of the amount of regularly reinforced training. On the other hand, it increased as a function of the amount of partial training up to 20 sessions and then leveled off at a value which clearly demonstrated the PRE. Resistance to extinction after training under the partial-regular sequence was never significantly lower than for partial alone. In fact, the addition of regular reinforcement increased resistance to extinction when the amount of prior training under partial was not sufficiently extensive to produce a maximum.

From these results and those of previous experiments the conclusion is well established that the abruptness of the local transition between training and extinction is not critically involved in the PRE.

References

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

HUMPHREYS, L. G. Extinction of conditioned psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–76.

JENKINS, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, **61**, 111–121.

JENSEN, G. D., & COTTON, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, **60**, 41–49.

KELLER, F. S. The effect of sequence of continuous and periodic reinforcement upon the "reflex reserve." *J. exp. Psychol.*, 1940, **27**, 559–565.

LAUER, D. W., & CARTERETTE, T. S. Changes in response measures over repeated acquisitions and extinctions of a running habit. *J. comp. physiol. Psychol.*, 1957, **50**, 334–338.

LAWRENCE, D. H., & FESTINGER, L. *Deterrents and reinforcements.* Stanford: Stanford Univer. Press, 1962.

LIKELY, F. A. Relative resistance to extinction of aperiodic and continuous reinforcement separately and in combination. *J. gen. Psychol.*, 1958, **58**, 165–187.

MOWRER, O. H., & JONES, H. M. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **34**, 293–311.

QUATERMAIN, D., & VAUGHAN, G. M. Effect of interpolating continuous reinforcement between partial training and extinction. *Psychol. Rep.*, 1961, **8**, 235–237.

THEIOS, J. M. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 1–6.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **48**, 318–322.

WEINSTOCK, S. Acquisition and extinction for a partially reinforced running response at a 24-hr. inter-trial interval. *J. exp. Psychol.*, 1958, **56**, 151–158.

(Early publication received May 7, 1962)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 451–459

PARTIAL REINFORCEMENT, CONTINUOUS REINFORCEMENT, AND REINFORCEMENT SHIFT EFFECTS[1]

STEWART H. HULSE

Johns Hopkins University

Hulse (1962) has shown that partial reinforcement (PRF) of running in a straight alley produces sharp discrimination of reward stimuli in the goal box, but continuous reinforcement (CRF) does not. Partially reinforced rats learned to lick from a drinking tube much faster than continuously reinforced rats, and the number of licks they emitted on nonreinforced running trials rapidly decreased to a very low level. Taken as a whole, the data indicated that behavior in the goal box was much more critically focused on the reward and its stimulus properties if PRF as compared with CRF were used.

There is an important implication of these experimental results. During the course of training with PRF, *changes* in the stimulus properties of the reward ought to be sharply discriminated. For example, on the assumptions that sweetness is a valid stimulus continuum for a reinforcing substance (Guttman, 1953) and response rate is positively correlated with sweetness, shifts from low to high concentrations of a saccharin reinforcer ought to produce prompt and direct increases in response rate. Discrimination of a stimulus change of this sort, and appropriate changes in response rate, ought to occur to a much lesser degree, if at all, with CRF. This would be true since CRF does not appear to bring behavior under the stimulus control of the reinforcer to the extent that PRF does.

The data reported here constitute a test of this implication. In one experiment, rats were conditioned on low and high concentrations of saccharin with PRF and CRF. Then, for half the rats, reinforcement concentrations were switched from low to high or from high to low. In a second experiment, essentially the same procedure was used, except that rats were switched only from low to high concentrations, and more extreme differences in concentration were used. Instrumentally conditioned licking was used as a response. This response is particularly appropriate, since the tongue is intimately involved in the sensory dimension of taste, and since licking rates vary quite consistently with the concentration of a sweet reinforcing stimulus (Hulse & Bacon, 1962).

EXPERIMENT I

Method

Subjects.—The *Ss* were 64 naive male albino rats, 85 days of age, of the Sprague-Dawley strain obtained from Sprague-Dawley, Incorporated, Madison, Wisconsin. The *Ss* had no experience drinking from tubes prior to their use in the experiment.

Apparatus and procedure.—The apparatus and the general procedure for treating licking as an instrumental response have been described in detail elsewhere (Hulse, 1960; Hulse, Snyder, & Bacon, 1960). In brief, the apparatus consisted of a 6.5 × 10 × 7 in. wooden box which was suitably lighted, ventilated, and sound-shielded. A piece of ¼-in. Plexiglas was located in front of a hole

[1] The research reported in this paper was supported by National Science Foundation Research Grants G8712 and G18125. The author is indebted to M. A. Berkley and E. J. Appelman for collecting the data.

in one wall of the box. The Plexiglas had a 1 × 2 cm. vertical slot cut into it. The *S* had access through the slot to a drinking tube.

Each lick on the drinking tube operated an electronic relay which operated, in turn, a programing circuit. Liquid reinforcements of a specified volume were delivered to *S* through the tube by means of an infusion pump operated by the programmer. For partial schedules of reinforcement, the programmer also operated a solenoid-driven lever. The lever slightly squeezed the pressure tubing leading from the pump to the drinking tube. After the reinforcement was delivered, the tubing returned to its normal shape and produced a slight negative pressure in the fluid system. The negative pressure drew the fluid column about 1 mm. inside the tip of the drinking tube. This assured that *S* could not make contact with the fluid on nonreinforced licks.

Two drinking tubes were used in the experiment. During the first 4 days of training, a brass tube with a 2-mm. fluid hole was used. On Day 1, this tube projected through the slot in the experimental box; on Days 2 to 4, the tube was gradually withdrawn through the slot. On Day 5, a plastic nipple with a 2-mm. fluid hole was introduced behind the slot and used for the rest of the experiment. This tube had a $\frac{1}{16}$-in. diameter brass electrical contact located just below the fluid hole.

The experimental design that was used during initial training incorporated two schedules of reinforcement (CRF and FR8), and two concentrations of reinforcement (1 gm. and 10 gm. of saccharin added to 1 l. of tap water). Sixteen *S*s were randomly assigned to each of the four groups called for by this design. After initial training, the concentrations of reinforcement were switched for half the *S*s trained with each schedule of reinforcement. For both the CRF and FR8 schedules of reinforcement, the postswitch design thus included one group switched from the low to the high concentration, one group switched from high to low, and two control groups which continued on their initial concentrations ($N = 8$ in each group).

Upon receipt from the breeder, *S*s were put in colony cages and maintained on ad lib. Purina lab chow and water (available from cans) for 10 days. Ten days of taming followed. On Day 1 of taming, *S*s were transferred to individual cages and placed on a daily deprivation diet of 10 gm. ground Purina lab chow mixed with 20 cc of tap water. They stayed on this deprivation schedule throughout the remainder of the experiment. Also, on each taming day, *S*s were taken in groups of approximately 15, freely handled, and allowed to explore in a large wooden box for 30 min. They received their daily ration of wet mash in their home cages after this procedure was completed.

Following taming, all *S*s received 22 days of initial conditioning of the instrumental licking response. On Day 1, each *S* was placed in the apparatus and permitted 300 licks from the brass tube on CRF. The concentration of reinforcement variable was introduced immediately. The size of each reinforcement was .005 cc of fluid. The plastic nipple was introduced on Day 5. The number of licks permitted all *S*s and the ratio of reinforcement for the partial *S*s gradually increased so that by Day 7, all *S*s made 1,000 licks, and by Day 18, the partial *S*s were on their final FR8. This regime was continued through Day 22.

On Day 23, the concentration of saccharin for *S*s in the switched groups was shifted from high to low or from low to high, depending upon the concentration that was used during initial training. This regime was continued from Day 23 through Day 29. On Day 30, *S*s in the switched groups were shifted back to their original concentrations and run under this regime through Day 36. The *S*s in the control groups remained on their initial concentrations throughout the 36 days of training.

On Day 37, 3 days of extinction began. Each day, *S*s were placed in the apparatus and allowed to lick the dry tube for 3 min. All conditions remained the same as those prevailing during training, except that the pump and solenoid-driven lever were disconnected from the programmer.

The *S*s were fed their daily ration of wet mash 20 to 30 min. after they had been run. The amount of water added to the mash was adjusted to account for the fluid obtained in the apparatus.

During training, the total time that each *S* required to emit its 1,000-lick allotment was recorded for each day. These times were transformed to reciprocals and multiplied by 1,000 to give a measure of average licking rate in licks per second. During extinction, the total number of licks that each *S* emitted during the daily 3-min. extinction period was recorded.

Results

Preswitch performance.—By Day 22, the last training day before the first switch in saccharin concentration,

CRF produced faster licking rates than PRF, and the high concentration of saccharin produced faster licking rates than the low concentration of saccharin. An analysis of variance for Day 22 for Ratio and Concentration of Reinforcement yielded an F for Ratio of 15.83 ($df = 1/60$, $P < .01$) and an F for Concentration of 59.11 ($df = 1/60$, $P < .01$). The interaction between the two variables was not significant ($P > .05$).

First switch.—The effect of the first switch in concentration, which occurred on Day 23, was quite different for CRF and PRF. Figure 1 shows that with PRF an orderly and appropriate change in rate occurred from Day 22 to Day 23. If the switch was from a high to a low concentration, licking rate decreased. If the switch was from low to high, licking rate increased. With CRF, on the other hand (Fig. 2), licking rates decreased regardless of the direction of the switch in concentration.

Statistical analyses of these results were carried out in two ways. First, analyses of variance were used on means for the groups that were switched in concentration (Table 1). These analyses were done separately for the CRF and PRF conditions, and they included pre- and postswitch days as a variable. Second, analyses of variance were used on means for both the control and switched groups (Tables 2 and 3). These were done separately for both the CRF and PRF conditions and for Days 22 and 23.

Table 1 shows a highly significant Treatments × Days interaction for the PRF condition. By t test, the licking rate of the PRF High-Switched group decreases from Day 22 to Day

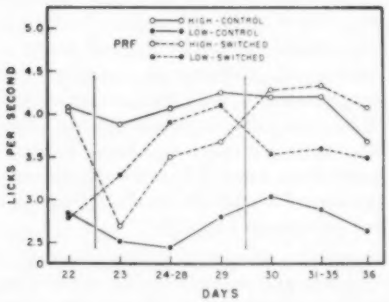

FIG. 1. Effect of shifts in concentration of reinforcement for PRF in Exp. I. (Shifts occurred between Days 22 and 23, and between Days 29 and 30.)

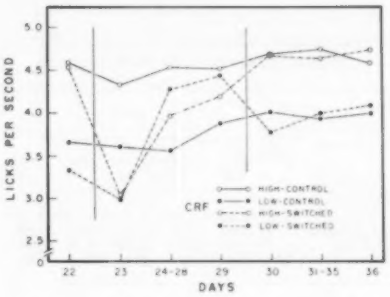

FIG. 2. Effect of shifts in concentration of reinforcement for CRF in Exp. I. (Shifts occurred between Days 22 and 23, and between Days 29 and 30.)

TABLE 1

ANALYSES OF VARIANCE OF SWITCHED-GROUP LICKING RATES DURING TRAINING

Source	*df*	*MS*			
		PRF		CRF	
		Days 22–23	Days 29–30	Days 22–23	Days 29–30
Ss	15				
Treatments (T)	1	0.98	0.21	1.67	0.84
Error	14	1.09	0.70	1.35	0.75
Days (D)	1	1.62	0.01	9.57**	0.05
Ss × D	15	0.67	0.30	0.56	0.26
T × D	1	7.22**	2.64**	1.32	2.65**
Error	14	0.20	0.13	0.51	0.09

** $P < .01$.

TABLE 2

ANALYSES OF VARIANCE OF LICKING RATES DURING TRAINING FOR PRF GROUPS

Source	*df*	*MS*				
		Day 22	Day 23	Day 29	Day 30	Day 36
Concentration (C)	1	12.50**	7.70**	6.57**	7.22**	5.36**
Switch (Sw)	1	0.01	0.38	0.94	0.67	3.06*
C × Sw	1	0.01	1.17	1.95*	0.36	0.43
Error	28	0.43	0.54	0.34	0.42	0.62

* $P < .05$.
** $P < .01$.

23 ($t = 6.36$, $df = 14$, $P < .01$), and the licking rate of the PRF Low-Switched group increases from Day 22 to Day 23 ($t = 2.27$, $df = 14$, $P < .05$). For the CRF condition, Table 1 shows that only Days is a significant variable. Licking rates on Day 23 are much lower than on Day 22, and this is true regardless of the direction of the switch in concentration. Tables 2 and 3 show that for both PRF and CRF on Day 22 only the Concentration variable is significant. High concentrations produced faster licking rates than low concentrations. On Day 23, however, the pattern of significant effects is quite different for PRF and CRF. For PRF, the Concentration variable once again produces the only significant effect. Licking rate is correlated with the concentration used on Day 23 for both switched and control groups. For CRF, only the Switch variable is significant. Licking rates are generally lower for the switched groups than they are for the control groups. By t tests, both the High-Switched and the Low-Switched groups differ from the High-Control group (t's = 2.72 and 2.83, $df = 28$, Ps $< .05$ and .01, respectively), but neither switched group differs significantly from the Low-Control group (Ps $> .05$).

TABLE 3

ANALYSES OF VARIANCE OF LICKING RATES DURING TRAINING FOR CRF GROUPS

Source	*df*	*MS*				
		Day 22	Day 23	Day 29	Day 30	Day 36
Concentration (C)	1	8.92**	0.92	1.62	4.97**	11.56**
Switch (Sw)	1	0.16	7.22**	0.10	0.13	0.12
C × Sw	1	0.31	1.20	0.32	0.09	0.02
Error	28	0.32	0.90	0.45	0.34	0.27

** $P < .01$.

Terminal performance on the concentrations of the first switch was quite different for PRF and CRF.[2] For PRF on Day 29 (Fig. 1), licking rates for the Low-Switched group reached about the same level as those of the High-Control group. Licking rates for the High-Switched group, however, showed an increase following Day 23 and reached an asymptote that was much higher than that of the Low-Control group. Table 2 shows significant effects for both Concentration and the interaction between Concentration and Switch in Concentration. By t test, the Low-Switched group does not differ from the High-Control group ($P > .05$), but the High-Switched group licks significantly faster than the Low-Control group ($t = 2.90$, $df = 28$, $P < .01$). The results for CRF (Fig. 2) show a pattern of differences among the groups which is similar to that for PRF, but the magnitudes of the differences are much smaller. Table 3 shows that for CRF none of the experimental variables produced significant effects on Day 29.

[2] Analyses of variance for the last 3 days of the first and second switches revealed no significant effects due to Days or the interaction of Days with any of the other variables. The data thus provide no evidence that the groups had not reached asymptotic performance by Day 29 or by Day 36.

Second switch.—With one major exception, the effect of the second switch in concentration on Day 30 was quite similar for both PRF and CRF. Figures 1 and 2 show that the licking rates of the switched groups changed in the appropriate direction when these groups returned to their original training concentrations. This was true regardless of ratio of reinforcement. The extent of the change in rate, however, was quite different for PRF and CRF. With CRF, Fig. 2 shows that the switched groups reached licking rates which were essentially the same as the control groups. With PRF, however, Fig. 1 shows that the Low-Switched group changed to a performance level which was considerably above that of the Low-Control group. By Day 36, the PRF High-Switched group also responded at a faster rate than the High-Control group.

Table 1 shows a significant Treatments × Days interaction for the switched groups for Days 29 and 30 for both PRF and CRF. Tables 2 and 3 show a significant Concentration effect on Day 30 for both PRF and CRF, but none of the other variables are significant. On Day 36, however, there is a significant Switch effect for PRF in addition to significant Concentration effects for both PRF and CRF. Although both PRF switched groups performed above their controls on Day 36, the difference between the Low-Switched and the Low-Control group is the only one that reaches significance ($t = 2.16$, $df = 28$, $P < .05$).

Extinction.—Fig 3 shows that, in general, the PRF groups emitted more licks on the dry drinking tube than the CRF groups. Moreover, with PRF, the switched groups were far more resistant to extinction than the nonswitched groups. With CRF,

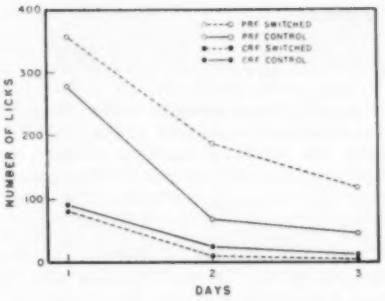

FIG. 3. Resistance to extinction in Exp. I for PRF and CRF Switched and Control groups.

however, the data suggest that the switched groups were less resistant to extinction than the nonswitched groups.

An analysis of variance on mean number of licks for all groups across the 3 days of extinction shows significant effects for Ratio of Reinforcement ($F = 70.48$, $df = 1/56$, $P < .01$), Days of Extinction ($F = 102.93$, $df = 2/126$, $P < .01$), Switch in Concentration ($F = 5.78$, $df = 1/56$, $P < .05$), Ratio × Switch ($F = 9.10$, $df = 1/56$, $P < .01$), and Ratio × Days ($F = 41.46$, $df = 2/112$, $P < .01$). For PRF, the mean difference between the switched and nonswitched groups is significant ($t = 6.61$, $df = 56$, $P < .01$). For CRF, this difference is not significant ($t = 0.76$, $df = 56$, $P > .05$).

EXPERIMENT II

Experiment II was run to check on the generality of some of the phenomena which resulted from the first switch of Exp. I. Experiment II incorporated more days of training before the switch took place and used more extreme differences in the concentration of the saccharin reinforcement.

Method

Subjects.—The *Ss* were 30 naive male albino rats, approximately 75 days of age, of the Sprague-Dawley strain.

Apparatus and procedure.—The apparatus was identical to that used in Exp. I. Similarly, the details of housing, taming, and feeding of *Ss* were essentially the same as those of Exp. I. A somewhat more severe daily deprivation regime was adopted: 9 gm. of Purina chow mixed with 18 cc of tap water.

The experimental design that was used was identical to that of Exp. I except that no High-Switched groups were run. The post-switch design thus included six groups with an N of 5 in each group. Reinforcement concentrations were either .50 gm. or 10 gm. of saccharin added to 1l. of tap water. As before, ratio of reinforcement was either CRF or FR8. The size of each reinforcement was .0043 cc of fluid.

Initial training in Exp. II was identical to that of Exp. I with the following exceptions. Only 600 licks were permitted each day. The PRF *Ss* were on their final FR8 schedule by Day 16, and all *Ss* continued initial training through Day 32. The PRF *Ss* thus received 17 days of training on FR8 instead of the 5 days that the PRF *Ss* received in Exp. I. Following initial training, concentration of reinforcement was changed from low to high on Day 33 for the PRF and CRF Switched groups, and the *Ss* continued on the new regime through Day 37.

Results

The immediate effects of the switch from a low to a high concentration are qualitatively exactly the same as those obtained in Exp. I. Figure 4 shows that with PRF, the licking rates of the Switched group increased markedly from Day 32 to Day 33. Comparison of the data shown in Fig. 4 with those shown in Fig. 1 suggests that the increase in rate was considerably greater under the conditions of Exp. II than under those of Exp. I. Figure 5 shows that with CRF, licking rates of the Switched group declined, as they did in Exp. I.

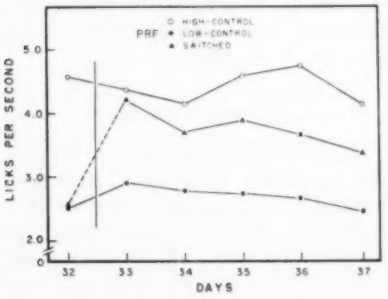

FIG. 4. Effect of shift in concentration of reinforcement for PRF in Exp. II. (The shift occurred between Days 32 and 33.)

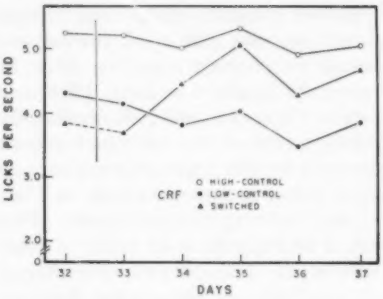

FIG. 5. Effect of shift in concentration of reinforcement for CRF in Exp. II. (The shift occurred between Days 32 and 33.)

An analysis of variance used on the groups means of Days 32 and 33 for PRF shows the following effects to be significant: Treatments ($F = 14.76$, $df = 2/12$, $P < .01$), Days ($F = 4.66$, $df = 1/14$, $P < .05$), and Treatments $\times$ Days ($F = 7.49$, $df = 2/12$, $P < .01$). The significant Treatments effect occurs primarily because of the large and consistent difference in performance between the High-Control and the Low-Control groups, while the significant Treatments $\times$ Days effect reflects the sharp change in performance of the Switched group from Day 32 to Day 33. An analysis of variance for CRF shows Treatments to be the only significant effect ($F = 12.61$, $df = 2/12$, $P < .01$). A third analysis of variance, which compared the performance of the

CRF and PRF Switched groups on Days 32 and 33, shows a significant Days × Ratio of Reinforcement interaction ($F = 9.14$, $df = 1/8$, $P < .05$). This corroborates the results shown by the separate analyses for PRF and CRF regarding the effects of the switch in concentration.

Following the first day of the switch in concentration, the performance of the CRF and PRF Switched groups was quite similar. Both groups reached a licking rate which was between the rates of the High-Control and Low-Control groups. An analysis of variance of means obtained by averaging rates on Days 36 and 37 shows a significant effect due to Treatments ($F = 18.86$, $df = 2/22$, $P < .01$) and to Ratio of Reinforcement ($F = 15.57$, $df = 1/22$, $P < .01$). The interaction between these two variables is not significant. By t test, the Switched groups responded significantly faster than the Low-Control groups ($t = 3.62$, $df = 22$, $P < .01$) and significantly slower than the High-Control groups ($t = 2.50$, $df = 22$, $P < .05$).

Discussion

The licking behavior that occurs on the day of a sudden change in the concentration of the reinforcer indicates that rats who are trained with PRF are conditioned to respond to the reinforcing stimulus in a way which is quite different from rats who are trained with CRF. With PRF, the sweetness and other stimulus properties of the reinforcer are sharply defined for *S*. This happens because of the contrast in stimulus conditions which occurs when responses are sometimes followed by a drop of saccharin and sometimes followed by nothing. If the concentration of the saccharin suddenly changes, the change is relatively easy to detect, and response rate immediately increases or decreases in the direction of the change. In effect, behavior comes under critical and orderly control of the stimulus properties of the reinforcer through discrimination training.

A different situation prevails when CRF is used. Here, each successive response is followed by the same sweet stimulus, and there is no stimulus contrast from one response to the next. With CRF, *S* learns to attack the drinking tube in a way which is vigorous, but relatively speaking, blind. If the concentration of the saccharin reinforcer changes, stimulus generalization decrement prevails, and behavior is markedly disrupted. Pure generalization decrement will hold the first time that *S* meets a sudden change in concentration, but not, of course, on later occasions. With the first change, *S* receives new information about reinforcement conditions and is, in fact, trained through contrast to discriminate somewhat more about the stimulus properties of the reinforcer. A second change in concentration comes as less of a surprise to *S*, and as Exp. I shows, response rates immediately increase or decrease appropriately.

The behavior that appears on the days which follow a change in reinforcement concentration, and the behavior that appears during extinction, indicates that experience with a particular set of reinforcement conditions on PRF produces residual effects which have a long-lasting influence on response strength. The discrimination training that PRF provides apparently yields a *learned* connection between licking behavior and a particular stimulus property of the reinforcer, such as its sweetness.

First, licking rates for the PRF switched groups reach postshift asymptotes which often lie between the licking rates that are characteristic of the high and low concentrations of saccharin. The *S*s behave much as if they were averaging the intensities of the pre- and postswitch reinforcing stimuli. Stimulus averaging of this sort seems to be most easily demonstrated when concentrations are changed from high to low (cf. Fig. 1). On the first postswitch day of the second

change in concentration of Exp. I, for example, the response rate of the PRF Low-Switched group falls sharply. But it remains well above the response rate of the Low-Control group. In Exp. II, a compromise in rate is also obtained when concentrations are changed from low to high. The PRF Switched group increases its licking rate on the first postswitch day, but then the rate falls to an asymptote which is again about half way between the rates of the control groups. Sometimes, stimulus averaging and comprises in response rate also appear with CRF. But when they do, they are far less easily obtained than they are with PRF. In Exp. I, there is no indication of a consistent compromise in response rate for CRF. However, in Exp. II, which involved more training days and larger differences in reinforcement concentration, a compromise in response rate is clearly obtained for CRF as well as for PRF. Apparently the more extreme conditions of Exp. II are required, however, before a compromise in rate will appear with CRF and instrumentally conditioned licking. In this connection, Premack and Hillix (1962) trained rats to lick 4% and 16% sucrose solutions from a conventional drinking tube, a procedure analogous to CRF conditioning as defined here, and then shifted the concentration of sucrose for these rats to 32%. The licking rates on the new concentration were consistently lower than the rate of rats maintained on 32% sucrose throughout the experiment. Apparently, experience with a particular reinforcing stimulus on CRF can make that stimulus have long-lasting effects on behavior under some conditions. The important point is that, for the present at least, these conditions appear less well defined for CRF than they do for PRF.

Second, if experience with a particular set of reinforcement conditions on PRF produces residual effects which permanently influence behavior, extinction should provide a sensitive test for these effects. This would be true since extinction removes *S* from direct exposure to the primary reinforcing stimulus. The data suggest that the control of behavior by the reinforcing stimulus is indeed permanent and that it does transfer to extinction. First, the extinction data of Exp. I show that a past history of shifts in the concentration of the reinforcer serves to markedly increase resistance to extinction when PRF is used. Further, Hulse (1958), Wagner (1961), and Hulse and Bacon (1962) have shown that with PRF resistance to extinction increases as amount of reinforcement increases. This is true whether amount of reinforcement is defined in terms of weight of food or sweetness of saccharin solutions.

None of these things are true when CRF is used. In Exp. I the CRF groups that had a past history of shifts in concentration of the reinforcer were, if anything, less resistant to extinction than the control groups. In other experiments which have used CRF and varied amount of reinforcement, the extinction results are at best unpredictable. Sometimes resistance to extinction increases as amount of reinforcement increases (Zeaman, 1949), sometimes it decreases (Armus, 1959; Hulse, 1958), and sometimes it does not vary with amount of reinforcement at all (Hulse & Bacon, 1962). Since *S* receives no discrimination training for the reinforcer with CRF, it is perhaps not surprising that these inconsistent results should be obtained. The data suggest that with CRF, *S* is stimulus bound. Response strength will be consistently correlated with some stimulus property of the reinforcer as long as *S* is directly exposed to it, but change the reinforcer—or remove it—and the correlation vanishes.

Finally, none of the data in either experiment reveals the slightest trace of elation, depression, positive contrast, or negative contrast (Pubols, 1960). The only time that a switched group licks either faster or slower than a control group, in a way which might suggest a contrast effect, occurs after the initial change in concentration of Exp. I. Here the rate of the CRF High-Switched

group drops below that of the Low-Control group, but it immediately bounces back to a point which is higher than the Low-Control group. As we have seen, it seems most meaningful in the context of the present experiments to view changes in rate of this sort as due to stimulus generalization decrement. In the same connection, it is interesting to note that most of the significant contrast effects that are reported in the literature are negative contrast effects (Spence, 1956; Pubols, 1960). Since all the earlier experiments have used CRF, perhaps generalization decrement was at least as important as some emotional-motivational factor in determining the contrast effects that were obtained.

Summary

Instrumentally conditioned licking was studied in two experiments as a function of ratio of reinforcement and shifts in the concentration of a saccharin reinforcer. With partial reinforcement (PRF), a shift in concentration produces an immediate change in licking rate in the direction of the shift. With continuous reinforcement (CRF), the immediate reaction to the shift is always a decrease in response rate. With PRF, asymptotic response rates reached several days after a shift are often compromises between the rates of control groups maintained on high and low concentrations. With CRF, a compromise of this sort appears only when relatively large amounts of training and relatively extreme differences in concentration of reinforcement are used.

The data suggest that, with PRF, behavior is more critically and permanently under the control of reinforcement stimuli than with CRF. This happens because PRF provides discrimination training for reinforcement stimuli, but CRF does not.

References

Armus, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

Guttman, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as a reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

Hulse, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

Hulse, S. H. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, **3**, 1–3.

Hulse, S. H. Discrimination of the reward in learning with partial and continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 227–233.

Hulse, S. H., & Bacon, W. E. Supplementary report: Partial reinforcement and amount of reinforcement as determinants of instrumental licking rates. *J. exp. Psychol.*, 1962, **63**, 214–215.

Hulse, S. H., Snyder, H. L., & Bacon, W. E. Instrumental licking behavior as a function of schedule, volume, and concentration of a saccharin reinforcer. *J. exp. Psychol.*, 1960, **60**, 359–364.

Premack, D., & Hillix, W. A. Evidence for shift effects in the consummatory response. *J. exp. Psychol.*, 1962, **63**, 284–288.

Pubols, B. H., Jr. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Wagner, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

Zeaman, D. Response latency as a function of amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Early publication received June 13, 1962)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 460–466

INFERENTIAL BEHAVIOR IN CHILDREN AS A FUNCTION OF AGE AND SUBGOAL CONSTANCY[1]

TRACY S. KENDLER AND HOWARD H. KENDLER

Barnard College *New York University*

Problem solution sometimes requires the integration of units of behavior that are already in the repertoire of *S* but have not previously been used in conjunction with one another. When the efficiency of integration increases gradually as a function of the number of trials the process is variously described as trial-and-error, instrumental, or selective learning. When an efficient solution occurs on the first trial, without any preceding trial-and-error behavior, the process is variously designated as reasoning, insight, or inference. The present experiment is one of a series that seeks to explore the relationship between these two processes in children (Kendler & Kendler, 1956; Kendler, Kendler, Pliskoff, & D'Amato, 1958; Kendler & Kendler, 1961). These studies have used an experimental paradigm, derived from Hull (1935, 1952), in which *S* is trained on three discrete behavior segments (A-B, X-Y, and B-G) and then presented with a test situation in which *S* is instructed to get G when only A and X are available. Problem solution requires the assembly of A-B and B-G.

The purposes of the present study were (*a*) to determine whether the ability to infer, as measured by these particular operations, increases with age and (*b*) to analyze the role of the integrating B stimulus in the inferential process.

[1] This research was sponsored by the National Science Foundation. The authors wish to express their appreciation for the cooperation extended by Orville Sipe, the principal of the Le Conte Public School, and his staff throughout the conduct of this research.

METHOD

Experimental design.—There were four experimental groups, 32 *S*s in each, arranged in a 2 × 2 factorial design. The main effects were age, kindergartners (K) vs. third-graders (3rd), and the constancy of training and test subgoals, constant (C) vs. switched (S).

Subjects.—The *S*s were 137 children drawn from the Le Conte Public School in Berkeley, California. Nine *S*s were eliminated, 5 of them because they made simultaneous A and X choices, and 4 due to inadvertent errors in experimental procedure. The data to be reported are based on the remaining 128 children, of whom 63 were boys and 65 girls. The kindergartners' mean age was 68.7 mo., range from 61 to 74. The third-graders' mean age was 103.9 mo., range from 96 to 115. Within each age level *S*s were randomly assigned to C and S groups without regard to age or sex.

Apparatus.—The portable aluminum apparatus used consisted of three distinct square panels 17.5 cm. on a side. Each panel, which could be exposed to *S*'s view singly or in combination with the others, corresponded to one habit segment. The center panel provided for the B-G segment. On its blue anodized surface was a circular opening 2.4 cm. in diameter into which *S* could drop the objects that served as subgoals in this experiment, namely, a glass marble with a 1.8-cm. diameter and a steel ball bearing with a 1.3-cm. diameter. If *S* dropped the correct subgoal into the circular opening, a small, shiny, gold fairy-tale or nursery-rhyme charm was propelled to a trough near the bottom of the panel. The incorrect subgoal delivered no reward. The charms, e.g., "little Bo Peep," "the cow that jumped over the moon," "the gingerbread man," etc., which were presented in a set sequence, served as the major goal (G).

The two side panels corresponded to the A-B and X-Y segments. The left panel was anodized pink and the right anodized yellow.

Each panel was equipped with a button which, when pressed, closed a circuit that led to the delivery of the appropriate subgoal (glass marble or ball bearing) to a trough near the bottom of that panel.

Procedure.—All *S*s were run individually and completed in one experimental session. The session began with the administration of the Peabody Picture Vocabulary Test (Dunn, 1959), which took from 5 to 15 min. and was followed immediately by training on the inference apparatus. One of the side panels was opened and *S* was told, "Press this button and see what happens." When the subgoal was delivered, *S* was instructed to pick it up, look at it, then return it to *E* so that he might have another turn. This side panel was closed, the other side panel opened, and the procedure repeated. After *S* had one forced trial on each panel, the procedure was repeated with the order of the sequences reversed. Thus at this point in training *S* had made two responses on the A-B and on the X-Y segments in an ABBA order. After these four forced trials, the doors of both side panels were opened and *S* was shown one of the subgoals and directed to, "Press the button that will get one like this." The procedure was repeated with each of the subgoals presented in an ABBA-BAAB order until the criterion of six successive correct responses was reached.

The next step of the training started with *E* opening the middle panel (after closing the two side ones) and directing the attention of *S* to a small window through which he could see a charm. He was told that it was a fairy-tale or nursery-rhyme charm which he would soon get an opportunity to examine closely. The aperture was pointed out and *S* was informed that if he dropped "the right thing" into that hole, the charm would drop into the tray. Then the marble and the ball bearing were placed in front of the B-G panel and *E* said, "One of these two little things is the right thing. Take one of them and drop it in the hole to see if it makes the charm come out." On the next trial *S* was again instructed to drop in the one that would make the charm come out. After that, training proceeded until a criterion of four successive correct responses was reached. *S*s who made two or less errors received no additional instructions while those who committed three or more errors were again urged to drop in the one that would make the charm come out. The relative positions of the marble and ball bearing (i.e., left and right) were varied in a random order *after each correct response* throughout the training on the B-G segment.

After preliminary training the test trial was introduced with the following instructions: "Would you like to see another charm? Very well, this time I won't put out any little things, but I will open all the doors. If you do what you are supposed to, you can make the charm come out. Go ahead." The *S* was allowed 60 sec. in which to make any response that he chose. If he did not press the A or X button during this time, *E* said, "Which button should you press to help you get the charm? Go ahead." After *S* pressed either button, he was allowed another 60 sec. to complete the sequence by dropping either subgoal into the B-G aperture before the trial was terminated. Thus all *S*s made either an A or X response and the trial was terminated either when *S* made a major goal response or when 60 sec. had elapsed since the subgoal had appeared.

For half of the children at each age level the subgoals were switched (S) between the preliminary training and the test trial so that if *S* made an A choice, he obtained the Y subgoal while if he made an X choice he obtained the B subgoal. For the remainder the subgoals remained in a constant position (C) during training and testing.

In order to minimize contamination from position and subgoal preferences, they were assigned in counterbalanced order: (*a*) Half of the *S*s in each experimental group had the right panel serve as the A-B segment while the left panel served as the X-Y segment. The opposite arrangement applied to the remaining half. (*b*) For half of the *S*s in each experimental group the marble was B and the ball bearing Y; for the remainder the ball bearing was B and the marble Y.

At the end of the experimental session each *S* was asked to keep the game a secret until everyone had a turn. It should be emphasized that at no point in the training or test did *E* describe the subgoal by name.

Results and Discussion

IQ scores.—The mean PPVT IQ scores for the four experimental groups were as follows: K-C (Kindergarten-Constant), 105.1; K-S, 106.8; 3-C (Third Grade-Constant), 110.5; and 3-S, 108.6. A 2×2 factorial analysis of variance applied to these data yielded no significant *F*s. The mean IQ for all *S*s was 107.8.

Preliminary training.—All groups quickly acquired the subgoal seg-

ments, i.e., they learned which panel yielded the marble and which yielded the ball bearing. Eighty-four percent of all *S*s attained criterion with no errors and 91% made no more than one error. No child made more than seven errors. The percentages of *S*s who attained criterion with no errors were as follows, by groups: K-C, 84%; K-S, 94%; 3-C, 75%; and 3-S, 78%. A χ^2 analysis of the corresponding frequencies revealed no statistically significant differences among the groups.

The major goal segment was also easily learned, i.e., whether the marble or the ball bearing yielded the charm. Except for 1 *S* who made nine errors and another who made four, all children reached the criterion after no more than three errors. The percentages of *S*s who attained criterion with no more than one error were as follows: K-C, 88%; K-S, 88%; 3-C, 69%; 3-S, 84%. A χ^2 analysis of the corresponding frequencies again yielded no significant differences among the groups. It should be noted, however, that in the acquisition of both the subgoal and major goal segments the older *S*s made more errors than the younger. This may be another manifestation of the response-shift-tendency reported by Harlow (1959).

Test trial: initial choice.—The first component of inferential behavior is the initial choice between A and X, in which A is the inferential choice. The combined results, which appear in Table 1, show a statistically significant age difference in the expected direction ($P < .01$). In fact, on this measure, the kindergarten children perform precisely at chance level.

TABLE 1

PERCENTAGE OF *S*S IN EACH EXPERIMENTAL GROUP WHOSE INITIAL CHOICE WAS A

Experimental Group	Chronological Age Level			
	5–6 Yr.		8–10 Yr.	
	%	σ_P	%	σ_P
Subgoals switched	50.0	.088	71.9	.079
Subgoals constant	50.0	.088	75.0	.077
Combined	50.0	.063	73.4	.056

This part of the results appears to be in conflict with two previous studies (Kendler & Kendler, 1956; Kendler et al., 1958) which showed that in groups of children between 34 and 60 mo. of age there were significantly more A than X choices. The studies, however, differed from the present one in two ways. The task used was simpler and the *S*s were from higher socioeconomic levels. A more recent study (Kendler & Kendler, 1961) used a procedure and socioeconomic sample comparable to the present ones and obtained similar results. In that study, which had as *S*s children between 30 and 65 mo. of age, the number of A and X choices was almost exactly equal. It is therefore suggested that the apparent conflict in results is due to the roles played by the difficulty of the task and the intelligence of the *S*s. To some extent this explanation of the discrepancy between the studies is supported by the findings of the present study (see Table 2) that the selection of A is related to *mental* age. It is evident that there was, in general, an increase in the percentage of A choices with increasing MA. This is, however, an ad hoc analysis, not tested for statistical significance; consequently the generality of the results is limited. They do, nevertheless, support the implication of the CA results, namely that inferential behavior is symptomatic of an important developmental process.

Test trial: integration response.—If after his initial choice, *S* inserted a subgoal into the G aperture, he was considered to have made an integration response. If the subgoal thus utilized was B, the integration re-

TABLE 2

PERCENTAGE OF A CHOICES AS A FUNCTION OF MA

MA	Chronological Age Levels					
	5–6 Yr.		8–10 Yr.		Both	
	% A	σ_P	% A	σ_P	% A	σ_P
36–71	40.0	.098	100.0	—	42.3	.097
72–95	48.3	.093	50.0	.145	48.8	.078
96–119	66.7	.192	75.0	.097	73.1	.087
120–143	100.0	—	71.4	.111	77.8	.098
144–167+	—	—	88.2	.079	88.2	.079

sponse was correct. If it was Y, the integration response was incorrect. Each of these integration responses could be further subdivided into those that occurred with no unnecessary responses intervening between initial and goal responses (direct) and those that occurred after one or more unnecessary responses (indirect). The intervening responses ranged from making only one unnecessary response, e.g., pressing the X button, to repeating almost the entire training sequence, e.g., pressing A but leaving B in trough, then pressing X and leaving Y in trough, then taking both subgoals out and setting them into the same position they occupied during training of B-G, then finally taking up B to drop into the G aperture. (The number of unnecessary responses in indirect solutions may be a fruitful subject for analysis in the future, but for the present the entire gamut was treated as an entity.) Finally, *S* had the option of making no integration response at all.

Table 3 presents the results for all *Ss* in the C groups, divided into these five categories. Since it might appear that the superiority of the older children of this measure is merely a reflection of the greater proportion of correct initial responses, the results were also analyzed using as a base only *Ss* who made an initial A choice, with the following effect. Twelve percent of the kindergartners and 67% of the third-graders who made an initial A choice made direct correct integration responses. This difference is significant ($P < .001$).

There are several conclusions to be drawn from the data in Table 3. One is that, as was the case with the initial choice measure, the integration measure yields marked differences in the performance of the two age levels. Note that the method used by those kindergartners who do attain a correct solution is primarily indirect. Descriptively speaking, such inference was more like trial-and-error than insightful behavior. On the other hand, a very large majority of the older *Ss* made a correct integration response, and their solutions tended to be more direct than indirect.

The other conclusions deal with the relations between the first and second components of the inferential sequence. Since many *Ss*, particularly among the kindergartners, who made an A choice either made no integration response at all or interposed other responses before completing the sequence, it may be concluded that the stimulus presented by the "B subgoal

TABLE 3

PERCENTAGE OF *Ss* AT EACH AGE LEVEL IN CONSTANT SUBGOALS GROUPS WHO MADE VARIOUS INTEGRATION RESPONSES

Chronol. Age (Yr.)	Kind of Integration				No Integration Response
	A-B-G+ (Correct)		X-Y-G− (Incorrect)		
	Direct	Indirect	Direct	Indirect	
5–6	6.2	43.8	6.2	6.2	37.5
8–10	50.0	37.5	3.1	3.1	6.2

in its trough" does not necessarily produce the B-G behavior segment. On the other hand, it is also clear that when an integration response does occur it is much more likely to be a B-G response than a Y-G response. This result could be due to the presence of the B stimulus since it was to this stimulus that the integration response was trained. However, in the Constant Subgoals groups the appearance of the B stimulus depends on an A response. The data have shown that, at least among the older *S*s, the A choice is not a matter of chance. It must be due to some problem solving activity, probably some covert system of responses. It is therefore equally possible that the B-G integration may be attributed to the same response mechanism that led to the correct initial choice.

In order to sort out whether the salient influence on the integration response is related to the correctness of the initial choice or to the correctness of the subgoal stimulus, the percentages of direct integration responses under four different conditions are presented in Table 4. Two of these conditions are drawn from the data of the Constant Subgoals groups reported above. The other two conditions are drawn from the Switched Subgoals groups whose second component results have not yet been reported. In this table R refers to the initial response and S to the subgoal stimulus.

When the proportions of direct integration responses, correct (+) and incorrect (−), under the various conditions are compared for the two age levels, it is apparent that there are too few *S*s in the kindergarten groups to yield any trends. The older *S*s' behavior, however, varies in an interesting and statistically significant way. A χ^2 analysis of the corresponding frequencies yields a P between .02 and .05. Third graders are most likely to integrate the two components when both their initial choice and the subgoal are correct. When one of these elements is incorrect, the probability of an integration response decreases; however, either element by itself does lead more than one-third of the older *S*s to integrate. When neither element is correct, the probability of integration is minimal.

In order to interpret these results it is necessary to examine more closely the nature of these elements. It is to be expected that S+ is more likely to produce integration than S−. But from Table 4 it can be seen that S+ by itself is not sufficient to explain

TABLE 4

PERCENTAGE OF *S*s AT EACH AGE LEVEL WHO MADE DIRECT INTEGRATION RESPONSES AS A FUNCTION OF THE CORRECTNESS OF THE INITIAL RESPONSE AND SUBGOAL

Subgoals	Condition Prior to Integration Response	Type of Integration	Age Level			
			5–6 Yr.		8–10 Yr.	
			N_T	% Integr.	N_T	% Integr.
Constant	R+S+	A-B-G+	16	12	24	67
Switched	R−S+	X-B-G+	16	3	9	44
Switched	R+S−	A-Y-G−	16	0	23	35
Constant	R−S−	X-Y-G−	16	12	8	12

integration at either age level. Among the younger *Ss* very little integration occurs even under S+ conditions. Among the older *Ss*, S+ is important but is not as adequate by itself as when in combination with R+. In fact R+ is so important that it can effect some integration even when it is combined with S−.

To understand why S+ by itself is not the adequate stimulus for integration, it must be realized that it does not stand for the B subgoal in isolation. It actually symbolizes a total stimulus complex of "B in a trough of one of the side panels." During training this particular stimulus compound was associated with picking up B and then returning it to *E*. There are quite a few *Ss*, particularly among the kindergartners, who, during the test situation, do just that. It is another stimulus compound, namely B and Y in front of the B-G panel, that is associated with picking up B and dropping it into the aperture. The data indicate that the generalization from the former stimulus compound to the latter is more likely to occur in 8–10 yr. olds than in 5–6 yr. olds. In the older age group it is more likely to occur after a correct initial choice than after an incorrect one.

When these results are represented in terms of S-R associations (Kendler & Kendler, 1962), it becomes possible to describe the psychological difference in the behavior of the two age groups as well as to suggest the relevant mechanisms for the inferential behavior reported in this study. There appear to be three important characteristics of this behavior. One is that making the correct initial choice in an inferential solution depends on the ability of *S* to generate a covert anticipatory response to the major goal and to respond appropriately to its cues. Such ability, according to this study, is positively related to age (within the limits tested). A second characteristic is that inferential solution depends on a "short-cutting" process in which some overt S-R sequences previously learned must drop out. The younger *Ss* had difficulty in this respect. Repeating some of the responses practiced during training prevented many from exhibiting direct inferential solutions. The third characteristic of inferential behavior is that it is not governed completely by external events. If it were, then it would be expected that the availability of B following a response to X should yield as much integration as when it follows a response to A. Since it does not, one can conclude that the integration response must be influenced by some internal stimulus component associated with the correct initial choice. This source of stimulation is sufficiently potent to compete successfully, in some cases, with external stimulation when there is a conflict between the two. This three-part characterization of inference is not offered as a theory. Its function is to analyze inferential behavior into more fundamental processes so that each may be investigated independently.

Summary

Children of two age levels, namely 5–6 and 8–10 yr., were presented with a task that required the linkage of two out of three discretely acquired segments of behavior. The solution consisted of making an initial choice between two of the segments (one correct and the other incorrect) and then integrating the product of that choice with the third segment. The solutions could be direct (inferential), i.e., the goal achieved without any unnecessary responses. They could be indirect, i.e., the goal achieved after the repetition of previously acquired but presently irrelevant behavior segments. The findings were: (*a*) Older *Ss* made significantly more correct initial choices than younger *Ss*. (*b*) About half of the younger *Ss* ultimately reached the goal, but their method was primarily indirect. Almost all of the older *Ss* achieved solution and a majority of them by direct inferential means. (*c*) In inferential solutions the integration of the subgoal and major segments was a joint function of the relevance of the external stimulation (as

determined by preliminary training) and the correctness of the initial choice. (*d*) When these two contributing factors were experimentally balanced against each other, it was found that the internal stimulation associated with the correct initial response was about as important a determiner of integration as the relevance of the external stimulus.

REFERENCES

DUNN, L. D. *Peabody Picture Vocabulary Test (PPVT)*. Nashville, Tenn.: American Guidance Service, 1959.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.

HULL, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solutions. *Psychol. Rev.*, 1935, **42**, 219–245.

HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

KENDLER, H. H., & KENDLER, T. S. Inferential behavior in preschool children. *J. exp. Psychol.*, 1956, **51**, 311–314.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem-solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KENDLER, H. H., KENDLER, T. S., PLISKOFF, S. S., & D'AMATO, M. F. Inferential behavior in children: I. The influence of reinforcement on incentive motivation. *J. exp. Psychol.*, 1958, **55**, 207–212.

KENDLER, T. S., & KENDLER, H. H. Inferential behavior in children: II. The influence of order of presentation. *J. exp. Psychol.*, 1961, **61**, 442–448.

(Received September 7, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 467–475

THE SEMANTIC MEDIATION OF EVALUATIVE MEANING [1]

FRANCIS J. DI VESTA AND DONALD O. STOVER

Syracuse University

Within the theoretical framework of Osgood's (Osgood, Suci, & Tannenbaum, 1957) theory of meaning the *sign* of an object is a primary symbol assumed to evoke a representational mediating response that is some part of the total behavior emitted by the organism when stimulated by the object itself. This response produces distinctive cues, mediating behavior that would otherwise not have occurred in the absence of previous association of the object with the word. The representational elements (r_m-s_m) correspond to the meaning of the sign, and because of the cue or stimulus components the sign can be conditioned to other stimuli initially lacking in meaning. Upon conditioning such symbolic stimuli are designated *assigns*. On the basis of the congruity principle distinctive responses may be produced by novel stimuli if the assign, within appropriate contextual arrangements, is used to label other novel stimuli (Osgood et al., 1957).

Typical studies of this process are those in which meaning of verbal stimuli have been changed by simple conditioning (Staats, Staats, & Biggs, 1958) and shown to generalize to synonyms of those stimuli (Staats, Staats, & Heard, 1959). Rhine and Silun (1958) demonstrated that development and strength of a concept-attitude were affected by the amount of reinforcement. Eisman (1955) provided reinforcements for a word and then associated that word with various colored objects, thereby strengthening the probability that the object would be chosen by the *S*. Using a Treatments × *Ss* design, Osipow (1960) provided evidence that a color-name associated with either a positive or negative evaluative word resulted in corresponding changes in preference for a nonsense figure when the color-name was associated with the figure. Since significant changes also occurred in the control group his results were inconclusive. Di Vesta (1962) compared the effects of reinforcing a neutral color-name with the effects of attaching the neutral color-name to a positive evaluative sign in the first stage of the mediation process. It was demonstrated that both procedures were effective in changing preferences for a nonsense figure when the color-name was subsequently associated with the figure.

The primary concern in the present study was to compare the effect of two different experimental designs commonly used in the studies summarized above. These experiments were intended to extend the findings of the previous investigations by testing two hypotheses: (*a*) that association of a neutral symbol with several signs, all of which represent similar polarities of the evaluative dimension of meaning, will result in movement in semantic space of the assign corresponding to the connotative meaning of the signs associated with it; and (*b*) that labeling a neutral stimulus object (nonsense figure) with the conditioned assign will result in an evaluation of that stimulus object corresponding to the acquired mean-

[1] This research was supported by Research Grant M-2900 from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service. Thanks are due to Donald L. Meyer for statistical assistance. Randall Martin aided in collecting and analyzing the data for Exp. II. The authors are greatly indebted to Paul A. Miller and David F. Sine of the Syracuse Board of Education; to Frank Liss, Principal of the Charles Andrews School; Elsie Platto, Principal of the Edward Smith School; Ruth O'Brien, Principal of the Sumner School; and to the teachers of the fifth grades for providing us with facilities and friendly assistance in working with the children involved in the study.

ing of the assign, independent of whatever effects may occur from acquired distinctiveness of cues resulting from labeling alone.

Several distinctions between the designs used in the present study and those used in previous investigations may be summarized: A Treatments × *Ss* design with repeated measurements on each *S* was employed in Exp. I, and a two-factor design with repeated measurements on each *S* was employed in Exp. II. Both experiments provide for control of learning-how-to-learn and warm-up effects. The effect of meaning was controlled by using experimental treatments in which neutral evaluative words were conditioned to the assign. In the Treatments × *Ss* experiment two additional controls were used, one for the use of a label without experimentally acquired meanings, and another for the evaluation of the stimulus object without the use of labels. And, finally, in both experiments the assign was associated with a nonsense figure as the stimulus object rather than with another verbal symbol.

Experiment I

Method

Design.—The general procedure used in the experiment was a modification of the mediated generalization paradigm. In Phase 1, the three sets of nonsense syllables were, respectively, conditioned to words with positive, neutral, and negative evaluative meanings. In Phase 2, the *Ss* learned to name each of three different nonsense figures with one of the three conditioned nonsense syllables, a fourth figure with an unconditioned nonsense syllable, and *Ss* experienced a fifth figure without naming it. Thus, one of the figures was associated with a positive-conditioned nonsense syllable (positive treatment), a second with a neutral-conditioned nonsense syllable (neutral treatment), a third with a negative-conditioned nonsense syllable (negative treatment), a fourth with an unconditioned nonsense syllable (control-UNS), and a fifth with no name attached to it (control). In Phase 3, the *Ss* rated each of the figures on three semantic differential scales.

Materials.—Three decks of 5 × 8 in. cards, with 24 cards in each deck, were used in the verbal conditioning phase. One positive, neutral, or negative evaluative word was printed in block letters 1¼ in. high on each card. Examples of the eight positive evaluative words are RIGHT, CLEAN, BRAVE, and SMART; examples of the eight neutral evaluative words are MIDDLE, USUAL, AVERAGE, and MEDIUM; and examples of the eight negative words are WRONG, WICKED, FILTHY, and STUPID. The negative and positive evaluative words were antonyms. Eight different words were used for each of the polar positions. The order in which the cards were arranged in the deck was determined at random with the restriction that each polar position was represented twice in every block of six words and that no more than one meaning occurred twice in succession in each block of six words. The same words were used in the second and third decks and differed from the first deck only in the order in which the words appeared.

Three sets of semantic differential rating scales were used. All scales were measures of evaluative meaning. The first was a practice set to help *Ss* understand the procedure. The Pretty-Ugly and the Wise-Foolish scales were used. The second set, comprised of the Cruel-Kind, and Wise-Foolish scales, was used to determine whether *Ss*' meanings for the nonsense syllables were changed in Phase 1. The third set, made up of Good-Bad, Pretty-Ugly, and Like-Dislike scales, was the measure used in Phase 3 to determine the evaluation of the figures. It should be noted that none of the scales duplicated any of the specific words used in the verbal conditioning procedure. All scales were five-point scales, with points 4 cm. apart. The *S* was permitted to check any place on the scale that he felt best represented his judgment. The intensity of *S*'s response was obtained by measuring the distance, from the extreme end of the scale to his check mark, in centimeters, 1 cm. represented extreme negative ratings, 9 cm. marked the midpoint of the scale, and 17 cm. represented extreme positive ratings.

Five nonsense figures were used, constructed according to the procedure described by Attneave and Arnoult (1956) as Method I for angular shapes with closed contours. Each figure was approximately 3½ × 4 in. and was made of heavy white oaktag paper painted gray. These figures were mounted on a white oaktag square with sides 5½ in. long.

Procedure.—The *S*s were first introduced to the task and to the method of using the semantic differential scales by having them rate three sets of two pictures (similar to those in the Stanford-Binet Intelligence Scales) on the Pretty-Ugly scale to determine their ability to make these evaluations. In addition, *S* rated examples of behavior such as "a boy crossing the street without looking," and fictitious characters such as "Donald Duck" on the Wise-Foolish scales. All *S*s proceeded succcessfully through this task.

A procedure similar to that described earlier by Di Vesta (1961) was used in Phase 1. The *S*s were instructed that they were to learn the meanings of three words (nonsense syllables) that they had not heard before. Nine nonsense syllables were used throughout the experiment, 1 *S* receiving any three at random from among these, for example, PID, LOM, and SUD might have been used for 1 *S* and CIT, POH, and LIM for another *S*. The polar position of evaluative meaning (positive, neutral, or negative) associated with a particular nonsense syllable was randomly arranged within the total experiment in order to balance evaluations that might have been based on associative characteristics of the syllables. (In Exp. II the syllables were found to have no more than chance effects.) The *S* was then given the first deck of cards and instructed to look at the top card and indicate which of the nonsense syllables it defined. If *S* thought, for example, that "plain" was a definition for LOM, he was to say "LOM is plain." The correction procedure was used. After *S* responded, *E* gave the correct response. If *S* was correct, he went on to the next card; if incorrect, he repeated the correct association before proceeding to the next card. This procedure was continued until *S* reached the criterion of responding correctly to two blocks of six words in each. This criterion was used since it signified that *S* responded correctly to each polarity at least four times with different words. If *S* did not reach this criterion after proceeding through the three decks twice, he was eliminated from the experiment. At the conclusion of this phase, *S* rated each of the nonsense syllables (assigns) on the Cruel-Kind and the Wise-Foolish rating scales.

In Phase 2 *S* was told that the syllables he had just learned about were also the names of certain figures. The *S* was shown five nonsense figures and told that since they were unlike any other figures that he may have seen, they could not be called by names like rectangles, triangles, and the like. The *S* was further instructed that four of the figures did have names which he was to learn, and that the names corresponded in three cases to the syllables he had just learned, while the fourth figure also had a name (the unconditioned nonsense syllable) but that he had not learned it, and that the fifth figure had no name. The *S* was then shown the figures, one at a time, in random order until he could name all labeled ones correctly in four successive presentations without error. A procedure similar to that used in the first stage was used in reinforcing correct responses. In order to randomize any effects of initial preference for the figure, pairing of assigns and figures was varied among *S*s.

In the third phase, *S* was asked to rate the figures, one at a time. The order in which figures were presented to a particular *S* was determined at random. Each figure was rated on three semantic differential scales, Good-Bad, Pretty-Ugly, and Like-Dislike.

In order to increase the level of motivation *S* was told that he would earn a toy if he participated in the experiment. At the conclusion of the session he was allowed to select his choice of trinkets from among crayons, baseball picture cards, balloons, and the like.

Subjects.—The *S*s were 24 children from one fifth-grade class in an elementary school. However, 4 *S*s were eliminated from the analysis because they failed to meet the criterion for learning in Phase 1 and 3 other *S*s were eliminated because they failed to use the rating scales appropriately. Thus, all analyses were based on an *N* of 17. The *S*s' CAs were 9 and 10 yr.

Results

An average of 42.70 (SD = 21.79) pairings, excluding criterial trials, of nonsense syllables with signs was required to learn the evaluative meanings of the assigns and an average of 28.25 (SD = 17.63) pairings was required to achieve the criterion in learning to associate the assign with the nonsense figures. The means and SDs for ratings of the nonsense syllables immediately after conditioning, on the Cruel-Kind and Wise-Foolish scales were compared.[2] Since no more

[2] Two tables in which are presented the means and SDs of assign ratings for each group in Exp. I and II, one table summarizing the learning data for Exp. II, and one table

than two scores overlapped in any one of the comparisons of the distributions of ratings for the three treatments, it was obvious that the differences were significant and no formal tests of significance were made.

Each *S* had been administered all treatments and used three scales in the final ratings of the figures within each treatment. Accordingly, a split split-plots analysis of variance was used.[3] A summary of the Treatments × Scales × *S*s analysis is presented in Table 1. As is evident from the table, the *F*s for the main effects of both Treatments and Scales were significant ($P < .01$). A test of the significance of the differences between the means of the scales indicated that the ratings made on the Pretty-Ugly scale tended to be generally more negative ($P < .01$) than the ratings made on either of the other two scales. (Similar results were found in Exp. II.) The ratings made on the Good-Bad and the Like-Dislike scales were not significantly different ($P > .05$). The difference in the Pretty-Ugly scale did not interact with treatments as indicated by the F (< 1.00) for Treatments × Scales interaction. Thus only treatment means obtained by combining the ratings of the three scales to obtain a total score were compared. The overall comparison of treatments using combined scores is represented by the F of 5.06 ($P < .01$).

The means and *SD*s of individual scale scores and total scores for Treat-

TABLE 1

ANALYSIS OF VARIANCE OF FINAL RATINGS OF FIGURES: EXP. I

Source	*df*	*MS*	*F*
Between *S*s	16	24.18	
Within *S*s			
Treatments (T)	4	195.62	5.06*
T × *S*s (Error 1)	64	38.70	
Scales (S)	2	63.83	7.24*
T × S	8	5.92	<1.00
S × *S*s (Error 2)	160	8.82	
Total	254		

* $P < .01$.

summarizing the analysis of variance testing the effects of nonsense syllables within each group in Exp. II, have been deposited with the American Documentation Institute. Order Document No. 7261 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[3] The assumption of homogeneity of variance was found to be tenable, via Bartlett's test, for all analyses of variance except where otherwise indicated.

TABLE 2

MEANS AND *SD*s OF RATINGS OF FIGURES ON EACH SCALE AND TOTAL SCALES FOR EACH TREATMENT: EXP. I

Treatment	Scales							
	Good-Bad		Pretty-Ugly		Like-Dislike		Total	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Positive	12.71	3.10	10.65	4.95	12.59	3.26	11.98	3.97
Neutral	9.76	4.24	7.59	3.57	9.29	4.17	8.88	4.11
Negative	7.00	5.01	6.00	4.78	7.06	5.42	6.69	5.10
Control-UNS	8.82	2.29	7.53	3.65	7.41	3.97	8.12	3.60
Control	8.94	3.84	7.59	3.73	10.06	3.90	8.86	3.96

ments are presented in Table 2. Duncan's multiple range test (Edwards, 1960) was used in testing the multiple comparisons between means. The Treatments × *Ss* error term was used in making the comparisons. The comparisons of the positive evaluative treatment with each of the four other treatments indicated that all differences were significant ($P < .001$). The comparisons of the negative evaluative treatment with the neutral evaluative treatment and the control without labeling were also significant ($P < .01$). The comparison of the negative evaluative treatment with the control in which only labeling was used was not significant ($.10 > P > .05$). None of the other comparisons among controls and the neutral treatment were significant ($P > .10$).

Experiment II

Method

Design.—The primary purpose of this experiment was to replicate a part of Exp. I while excluding carry-over effects from one treatment to the next, a consideration often neglected by previous investigators. Nine groups, in a two-factor design with three repeated measurements in each group, were used. One factor was based on the three polarities (positive, neutral, and negative) of evaluative meaning; the second was based on the use of three different figures. Three figures were used to increase the generality of the results. If no differences were found between figures, the results could then be attributed more definitely to the mediation of meaning. In Phase 1 the nonsense syllable was associated with signs having positive-evaluative meanings for the positive group; with neutral-evaluative meanings for the neutral group; and with negative-evaluative meanings for the negative group. In Phase 2 the conditioned nonsense syllable was associated with Fig. 1 by one-third of each group, Fig. 2 by another third, and Fig. 3 by the last third of each group. In Phase 3 *Ss* used the three scales, as in the previous experiment, for rating the figure labeled by the assign.

Procedure.—The following modifications were made in the procedure used in Exp. I. In Phase 1, *S* learned the meaning for only one nonsense syllable. The same list of words was used as in Exp. I and in exactly the same form. However, in Exp. II *S* indicated whether or not each of the 24 words was the meaning for the nonsense syllable to be conditioned. Thus, for example, if he was in the neutral group he would say, "LOM is plain" or "LOM is not brave." The *E* then recited the correct association. If *S* was correct he turned to the next card, if incorrect he repeated the correct combination and then turned to the next card. Eight of the 24 words, depending upon the experimental condition, were used as the definition of the nonsense syllable in each group. In Phase 2 *S* then identified and named only three figures; one of the figures was labeled by the conditioned nonsense syllable and the other two figures were named with unconditioned nonsense syllables. The three figures were selected on the basis of pretest data with over 150 children in which these figures had been ranked as the most neutral of six figures.

Three different nonsense syllables were used in each group. Different syllables were used between groups. The fact that the three syllables might have comprised a fourth factor in the total design was inadvertently neglected. However, since equal numbers of *Ss* were assigned each of the syllables within any one condition, the data were analyzed by separate analyses of variance to determine whether the syllables used within each treatment had other than chance effects on the results. In the three analyses thus made, figures and syllables were fixed factors and scales were repeated measures. In every case, the main effect of syllables and the interaction term in which syllables appeared were not significant ($P > .20$). All but 2 of the 12 *F*s so calculated were < 1.00.[2] Accordingly, the conclusion was that the selected syllables had no effect on the final ratings made.

In Phase 3, the *S* rated the figure labeled with the conditioned syllable. The ratings were made with the same three scales used in Exp. I. The figures labeled with the unconditioned nonsense syllables were not rated since these data were not relevant to the hypotheses.

Subjects.—The *Ss* were 81 children from the fifth grade classes of a different elementary school from that used in Exp. I. The *Ss* were 9 and 10 yr. of age, and were randomly assigned to the nine experimental groups. Twelve *Ss* had been eliminated; 6 *Ss* failed to learn the task in Phase 1, and 1 *S* in the positive group, 4 *Ss* in the neutral group, and

1 in the negative group failed to use the rating scales properly at the end of the experiment.

Results

The means and *SD*s of the learning data as measured by errors and trials to reach the criterion for the three main experimental groups in Phases 1 and 2 were compared.[2] The variances of the learning data in Phase 1 were heterogeneous requiring the use of the Kruskal-Wallis analysis of variance to test differences among the groups. Significant ($P < .01$) differences among groups were found for both measures. Multiple comparisons were tested by the Mann-Whitney *U* test. The difference in errors between the positive and negative group was not significant ($P > .05$) while the comparisons of each of those groups with the neutral group were significant ($P < .01$). All comparisons on the "trials" criterion were significant ($P < .01$). However, the order of difficulty in learning was not correlated with the order of the final preference rating; therefore, the assumption was made that trials to learn or errors made in learning were not related to changes in preferences, and that any differences in final preference could be attributed to the mediation of meaning. No significant differences were found in the comparisons based on either of the learning measures in Phase 2.

The means and *SD*s of the assign ratings following the verbal conditioning procedure in Phase 1 were summarized.[2] These data compared favorably with those obtained for Exp. I.

The Treatments × Figures × Scales analysis of variance of the ratings made of the figures in Phase 3 is summarized in Table 3. While the overall Treatment effects were significant ($P < .001$) the differences in Scales ($P < .001$) and the interaction of Treatments and Scales ($P < .01$) were also significant. The means and *SD*s of the ratings of the figures on each of the three scales are presented in Table 4. The overall ratings are also summarized to permit comparison with the data from Exp. I. Data for the specific figures have been combined since the *F* for differences between figures was < 1.00. In the multiple comparisons[4] all scales significantly ($P < .01$) differentiated the positive from the negative group; the Good-Bad and the Like-Dislike scales significantly ($P < .05$) differentiated the neutral from the negative group; and the Pretty-Ugly scale significantly ($P < .05$) differentiated the neutral from the positive group. None of the other comparisons was significant. When the data from the three scales were combined the differentiation among the three groups was more clear-cut and the trends clearly corresponded to the conditioning of the syllables as well as to the data for

TABLE 3

ANALYSIS OF VARIANCE OF FINAL RATINGS OF FIGURES: EXP. II

Source	*df*	*MS*	*F*
Between *Ss*	80		
Treatment (T)	2	382.86	10.05**
Figures (F)	2	6.33	<1.00
T × F	4	12.53	<1.00
Error (1)	72	38.10	
Within *Ss*	162		
Scales (S)	2	78.48	8.79**
T × S	4	32.98	3.69*
F × S	4	18.30	2.05
T × F × S	8	12.66	1.42
Error (2)	144	8.93	
Total	242		

* $P < .01$.
** $P < .001$.

[4] The modification of Duncan's multiple comparison test described by Collier (1958) was used in testing the differences between treatment means on each scale.

TABLE 4

FINAL RATINGS OF FIGURES ON EACH SCALE AND TOTAL SCALES BY EXPERIMENTAL GROUPS: EXP. II

Group	Scales							
	Good-Bad		Pretty-Ugly		Like-Dislike		Total	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Positive	10.67	4.01	8.78	3.75	11.00	4.64	10.15	3.26
Neutral	9.30	3.70	6.96	3.40	9.89	4.10	8.59	3.27
Negative	4.52	4.80	6.19	4.60	6.93	4.80	5.88	3.87

Exp. I. The difference between the neutral and positive groups were significant ($P < .05$). The differences in the remaining two comparisons were also significant ($P < .05$).

DISCUSSION

The results of the two experiments demonstrate that (*a*) assigns acquire evaluative meaning through association with several signs having specific polarity on this dimension of meaning and that (*b*) these assigns transfer to affect the ratings of figures associated with them. The changes in preferences for the figures corresponded to the meanings, rather than the specific words, with which the nonsense syllable had been associated. Where such associations had not been made, as in the two control groups of Exp. I, the result was a neutral evaluation of the figure. These findings imply the occurrence of the mediation chain consisting of figure, nonsense syllable, associations, and connotative meaning. However, it is evident that the procedures used impose a severe limitation on this interpretation making an alternative explanation based on simple association principles an equally likely assumption. Thus, once the assign had been formed in the first learning phase it became a "synonym" for the associated words. These words, in turn, may have been directly or implicitly associated with the figure during the second learning phase when the figure was associated with the nonsense syllable. A more convincing demonstration of the mediation process would necessitate a comparison of the present experiments with those in which the two learning phases are reversed that is, a comparison of the present experiments with one in which the standard mediation paradigm is employed. These studies are currently being planned and are to be conducted in exactly the same settings in which the present studies were made.

Since the main concern, however, was in a comparison of the two designs, their respective merits and limitations warrant attention. Basically, they yielded the same results and one is, in a real sense, a replication of the other. The intra-individual design appears to be more sensitive in the gross demonstration of the effect than is the individual treatments design. Among the assumptions of the intraindividual design are that there are no order effects; and that there are no carry-over effects from one treatment to another. The first is tenable on the premise that treatments were assigned in random order to *S*s. The second, however, is a very likely occurrence in the major phases of Exp. I where *S*s learned the three meanings concomitantly. The comparisons between the stimuli conceivably facilitate the distinctions in evaluative response. Although the order of the figure was presented at random, once one figure had been rated a comparison with the others was inevitable; and, since the same three scales were employed in each rating it would also be expected that carry-over effects would occur in the use of the scale.

In both experiments, there were significant differences among the results obtained from the different scales. In all but one instance, that is, in the use of the Pretty-Ugly scale in Exp. II, the preferences for all figures based on each of the scales were in the predicted direction. The total score provided more adequate discrimination between groups in both experiments. It appears that the primary factor in the slight differences among scales may be accounted for in terms of lowered reliability when only one scale is used. The fact that other investigators, e.g., Staats, Staats, & Biggs (1958), have used a single scale with reliable results may merely reflect the ability of adults to use the scales with greater accuracy than children.

Within the limitations discussed above, the present study tentatively suggests the applicability of the representational mediation process to the development of attitudes (Dodge, 1955; Osgood et al., 1957). In terms of the operations of measurement with the semantic differential the meaning of a stimulus object is its allocation in multidimensional space, with attitude defined as the projection of this point onto the evaluative dimension of that space (Osgood et al., 1957). The principal concern is with the common elements of the mediating response of the meanings of the signs. When the assign is associated with evaluative signs, the meaning of the assign will depend upon the mediating response evoked by these signs. The assign may be attached to other neutral stimulus objects to influence the responses evoked by these stimuli. Attitudes are thus characterized as implicitly learned processes with potentially bipolar evaluative properties.

Summary

This study compared two experimental designs using procedures for the study of the mediation process. In both experiments the first phase consisted of assign development by conditioning signs with evaluative meaning to neutral nonsense syllables. Phase 2 involved associating the assign with a neutral nonsense figure. In the third phase *Ss* rated the nonsense figure on three semantic differential evaluative scales.

There were 17 fifth-grade children in Exp. I. Each *S* received five treatments. The dependent variable was the rating of five different nonsense figures. Three of the figures were labeled by assigns previously conditioned in Phase 1 to signs having negative, neutral, or positive evaluative meanings, respectively. A fourth figure was labeled by an unconditioned nonsense syllable and the fifth figure was experienced equally often with the others but was not labeled. Ratings, made on each of the scales, of the figures labeled with conditioned assigns corresponded with the evaluative meaning of signs associated with the assigns. No significant differences were found between ratings made of the figure labeled by the neutral assigns, that labeled by the unconditioned nonsense syllable, or the figure not labeled.

In Exp. II there were 81 fifth-grade pupils. In a two-factor design with repeated measures, independent groups were required to associate signs, varying, respectively, on the negative, neutral, and positive polarities of evaluative meaning, to assigns. Within each group the conditioned assigns were used to label one of three different nonsense figures. The remaining two figures were labeled by nonsense syllables. The results of this experiment corroborated those of the first. The overall ratings, using the combined scores of the three scales, differentiated significantly between the three groups. There was, however, a significant interaction between treatments and scales, although in all but one comparison the differences were in the predicted direction. No significant differences were found between figures. Nor were there significant differences between the nonsense syllables used with any one of the treatment groups.

References

Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **63**, 452–471.

Collier, R. O. Analysis of variance for correlated observations. *Psychometrika*, 1958, **53**, 223–236.

Di Vesta, F. J. Contrast effects in the verbal conditioning of meaning. *J. exp. Psychol.*, 1961, **62**, 535–544.

Di Vesta, F. J. The effects of mediated generalization on the development of children's preferences for figures. *Child Develpm.*, 1962, **33**, 209–222.

DODGE, J. S. A quantitative investigation of the relation between meaning development and context. Unpublished doctoral dissertation, University of Illinois, 1955.

EDWARDS, A. L. *Experimental design in psychological research.* N. Y.: Rinehart, 1960.

EISMAN, B. S. Attitude formation: The development of a color-preference response through mediated generalization. *J. abnorm. soc. Psychol.*, 1955, **50,** 321–326.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

OSIPOW, S. H. The effects of verbal mediation on the modification of children's attitudes. *J. educ. Psychol.*, 1960, **51,** 199–207.

RHINE, R. J., & SILUN, B. A. Acquisition and change of a concept-attitude as a function of consistency of reinforcement. *J. exp. Psychol.*, 1958, **55,** 524–529.

STAATS, A. W., STAATS, C. K., & BIGGS, D. A. Meaning of verbal stimuli changed by conditioning. *Amer. J. Psychol.*, 1958, **71,** 429–431.

STAATS, A. W., STAATS, C. K., & HEARD, W. G. Language conditioning of meaning using a semantic generalization paradigm. *J. exp. Psychol.*, 1959, **57,** 187–192.

(Received September 13, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 476–481

REVERSAL AND NONREVERSAL SHIFTS IN CONCEPT FORMATION USING CONSISTENT AND INCONSISTENT RESPONSES

MARTIN HARROW[1] AND ALEXANDER M. BUCHWALD

Indiana University

Several previous card-sorting and block-sorting experiments have compared the speed of making reversal and nonreversal shifts (Buss, 1953, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Mayzner, 1956; Kendler & Kendler, 1959). Typically, a reversal shift has involved learning two successive sorting tasks, with correct responses for Task 2 being based on the same dimension of the stimuli as on Task 1, but with *S* being required, literally, to reverse his previous sorting responses. For example, in Task 1, red cards must be placed in Sorting Category A and green cards in Category B, and in Task 2 red cards belong in Category B and green cards in Category A. Nonreversal shifts have involved learning one sorting task and then learning a second sorting task based on some dimension of the stimuli which was irrelevant in Task 1. Previous sorting experiments with college students have found that reversal shifts are learned quicker than nonreversal shifts.

In a number of the experiments in this area (Buss, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Kendler, 1959) it has been hypothesized that reversal shifts are learned more quickly because *S*s respond to the same dimension of the stimuli as was used previously in Task 1 learning, whereas in a nonreversal shift they are required to respond to a new dimension of the stimuli. According to this analysis, reversal shifts are learned quickly because the same cues that were used during Task 1 learning are again relevant during the learning of Task 2. The *S* must merely learn to make different responses to the previously used cues. On the basis of this hypothesis the results in the above experiments have been interpreted as supporting a mediational S-R framework. However, reversal groups in the above experiments, besides using the same dimension of the stimuli that was previously relevant, have also been required to make literal reversals, in which the exact opposite sorting response was required.

If the analysis in terms of the advantages of using the same dimension of the stimuli is correct, then this condition alone should be sufficient to produce facilitating effects. According to the above analysis supporting the mediational S-R framework, it would be expected that a shift which allows *S* to respond to previously relevant cues (for purposes of simplification and to maintain the traditional terminology used previously, this type of shift will also be called a reversal shift) would be learned quickly even if the new responses that are required are not the exact opposite of the previously learned responses. Thus, after learning a concept based on the *number* of stimulus elements, a reversal group switching to a second *number* concept should learn more quickly than a nonreversal group which shifts to a concept based on the

[1] Now at Yale University School of Medicine.

position of the stimuli, regardless of whether or not exactly opposite sorting responses are required. The present experiment attempted to test this analysis.

An interesting factor which has been involved, in an incidental manner, in some of the previous experiments concerning reversal and nonreversal shifts (Kendler & D'Amato, 1955; Kendler & Mayzner, 1956) was also investigated. This variable is the use of consistent and inconsistent responses. These responses occur in experiments which require *Ss* to sort response cards into compartments labeled with stimulus cards. Consistent responses occur when response cards are placed with stimulus cards which are similar to them in some respect (e.g., when *S* is required to sort response cards according to their color and must place red response cards into a compartment labeled with a red stimulus card). Inconsistent responses occur when there is little or no similarity between the stimulus card and its appropriate response cards (e.g., when *S* is required to sort response cards according to their color, but must put red response cards with a nonred stimulus card).

While consistent and inconsistent responses have appeared in some of the experiments comparing reversal and nonreversal shifts the influence of this variable on the ease of making these shifts has not been directly investigated. It is possible that the relative difficulty of reversal shifts as opposed to nonreversal shifts may be affected by whether the responses required are consistent or inconsistent. The present experiment was designed to investigate the influence of this variable. This was done in a card-sorting situation which varied consistent and inconsistent responses in a systematic fashion while comparing reversal and nonreversal shifts.

Method

Subjects.—The *Ss* were 83 undergraduate students enrolled in introductory psychology courses at Indiana University. Nineteen of these *Ss* were eliminated for failure to learn the first concept.

Materials.—The experiment was carried out in a four-category card-sorting situation in which response cards were sorted into categories labeled with stimulus cards. A deck of 16 response cards, 3 × 3 in. in size and similar in principle to the Wisconsin Card Sorting Test (Berg, 1948) was devised. Small green stars were placed on these cards and the cards varied according to the number and position of the stars on them. There were one, two, three, or four stars placed on each response card. The stars were arranged systematically with all possible combinations of number of stars appearing in each of the four corners of the cards. This made a set of 16 response cards, each of which could be sorted according to the number of stars on it or according to the position of the stars on it.

There were 4 stimulus cards and the *Ss* were required to sort the 16 response cards with them. These stimulus cards were attached to a black card holder. The first stimulus card had one black line in the upper-left hand corner (Card UL1), the second card had two black lines in the lower-left-hand corner (Card LL2), the third had three black lines in the lower-right hand corner (Card LR3), and the fourth had four black lines in the upper-right hand corner (Card UR4). At the top of the card holder, which the stimulus cards were attached to, was a pocket into which could be placed a response card.

Three different "position" concepts were used (where the correct sorting of each response card depends on the position that the stars are placed on the card) and three different "number" concepts were used (where the correct sorting of each response card depends on the number of stars on the card). The three position concepts were a consistent position concept and two inconsistent position concepts (these latter two were a clockwise position concept and a counterclockwise position concept). The consistent position concept required that response cards which had stars located in Corner X be placed with the stimulus card which had its lines located in the *same* corner, Corner X. For example, response cards having stars in the upper-left-

hand corner were placed with Stimulus Card UL1, which also had its lines located in the upper-left-hand corner. Likewise response cards having stars in the lower-left-hand corner belonged with Stimulus Card LL2, response cards having stars in the lower-right-hand corner went with Stimulus Card LR3, and response cards having stars in the upper-right-hand corner belonged with Stimulus Card UR4. One of the inconsistent position concepts (the clockwise concept) required that response cards which had stars in Corner X be sorted with the stimulus card which had its lines located in the corner which is one position clockwise to Corner X. Thus, response cards having their stars located in the lower-left-hand corner had to be sorted with Stimulus Card UL1, which had its lines located in the upper-left-hand corner (one position clockwise). Likewise, response cards having stars in the lower-right-hand corner belonged with Stimulus Card LL2, response cards having stars in the upper-right-hand corner went with Stimulus Card LR3, and cards having stars in the upper-left-hand corner belonged with Stimulus Card UR4. The other inconsistent position concept (the counterclockwise concept) required that response cards which had stars in Corner X be sorted with the stimulus card which had its lines located in the corner which was one position counterclockwise to Corner X.

The three "number" concepts were a consistent number concept, and two inconsistent number concepts—an $(n + 1)$ concept and an $(n - 1)$ concept. The consistent number concept required that response cards with n stars on them be sorted with the stimulus card which also had n lines on it. Thus, response cards which had one star had to be sorted with the stimulus card with one line (Card UL1). Likewise, response cards having two stars belonged with Stimulus Card LL2, response cards having three stars went with Stimulus Card LR3, and response cards having four stars went with Card UR4. One of the inconsistent number concepts—the $(n + 1)$ concept—required that response cards which had n stars be sorted with the stimulus card which had $(n + 1)$ lines. Hence, response cards which had one star had to be sorted with the stimulus card which had two lines (Card LL2). Similarly, response cards having two stars belonged with Stimulus Card LR3, response cards having three stars went with Stimulus Card UR4, and response cards having four stars belonged with Stimulus Card UL1. The other inconsistent number concept—the $(n - 1)$ concept—required that *Ss* sort response cards which had n stars with the stimulus card which had $(n - 1)$ lines on it.

Design.—In the present experiment an attempt was made to test the previous analysis that the facilitative effects of reversal shifts are due to the advantages of using the same dimension of the stimuli when shifting concepts. An attempt also was made to examine the influence of consistent and inconsistent responses on reversal and nonreversal shifts. This was achieved in a three-dimensional factorial design, in which *Ss* learned two successive card-sorting tasks. During Task 1 all correct responses were inconsistent with half of the *Ss* learning an inconsistent position concept and the other half learning an inconsistent number concept.

For Task 2 *Ss* were divided into eight subgroups with half of the *Ss* learning a concept involving the same dimension of the stimuli as was previously relevant (reversal shifts) and half learning a concept involving a different dimension of the stimuli than was previously relevant (nonreversal shifts). Half of the *Ss* learned a concept involving consistent responses and half learned a concept involving inconsistent responses. Similarly, half of the *Ss* learned position concepts and half learned number concepts. This permitted a $2 \times 2 \times 2$ factorial design.

Procedure.—The *Ss* were tested individually. The instructions read to *S* indicated that as each response card was shown to him he should point to the stimulus card he thought it belonged with, and that by *E* telling him whether he was right or wrong he would gradually find out where each response card really belonged.

The response cards were presented randomly and individually to each *S* by being placed in the slot which was at the top of the card holder. The criterion of learning for both Tasks 1 and 2 was 12 successive correct responses. All *Ss* who met the criterion on Task 1 within 160 trials were required to learn Task 2.

The change in the pattern of reinforcement for the second concept was made without informing *Ss*. The *Ss* who had not learned the second concept within 500 trials were arbitrarily assigned a score of 500 for Task 2 learning.

In order to avoid partial reinforcement of the first concept during learning of the second concept (Buss, 1956; Gormezano & Grant, 1958; Harrow & Friedman, 1958) 4 response cards were eliminated from each set of 16 cards, for each group of nonreversal *Ss*. This left 12 response cards in each deck. The same

TABLE 1

NUMBER OF TRIALS TO LEARN FIRST CONCEPT (TASK 1) AND SECOND CONCEPT (TASK 2)

Condition					N	Trials to Learn					
Task 1		Task 2		Type of Shift[b]		Task 1			Task 2		
Response[a]	Concept	Response[a]	Concept			Mean	Mdn.	Range	Mean	Mdn.	Range
Incon.	Position	Incon.	Position	Rev.	8	79	84.5	13–133	11	5.0	2–36
Incon.	Position	Incon.	Number	NR	8	55	44.5	6–136	98	46.0	20–231
Incon.	Position	Consist.	Position	Rev.	8	62	60.0	10–127	9	9.0	3–18
Incon.	Position	Consist.	Number	NR	8	71	65.5	40–133	18	14.0	2–50
Incon.	Number	Incon.	Number	Rev.	8	47	38.5	25–91	14	14.5	2–30
Incon.	Number	Incon.	Position	NR	8	25	24.5	3–44	155	107.0	8–488
Incon.	Number	Consist.	Number	Rev.	8	40	28.0	3–110	6	4.5	1–17
Incon.	Number	Consist.	Position	NR	8	53	55.0	2–100	38	15.0	4–178

[a] Incon. = inconsistent; consist. = consistent.
[b] Rev. = reversal; NR = nonreversal.

4 response cards were also eliminated for the corresponding reversal groups.

RESULTS

The data concerned with the learning of the first concept are reported in Table 1. Both these results and those for the second concept represent the number of trials to learn the task, excluding the 12 criterion trials. To test for differences in speed of learning between *Ss* of different groups learning the same first task, two separate analyses of variances were computed. One analysis of variance compared the four groups initially learning an inconsistent position concept and one compared the four groups initially learning the inconsistent number concept. Using 3 and 28 *df* in each case for the position and number concepts, the overall *F*s were, respectively, 0.48 ($P > .05$), and 1.36 ($P > .05$). Thus, the data indicate that the four groups of *Ss* learning each type of task were equated with each other initially.

The results for learning the second concept are also presented in Table 1. In order to determine whether there were any differences in the speed of learning the concepts among the eight experimental groups, a $2 \times 2 \times 2$ analysis of variance was carried out. Due to the skewness of the data, a logarithmic transformation was used in place of the raw scores to obtain homogeneity of variance (Edwards, 1950). The results, as can be seen in Table 2, show that the reversal groups learned significantly faster than the nonreversal groups ($P < .001$). Likewise, concepts requiring consistent responses were learned significantly faster than those requiring inconsistent responses ($P < .001$). The sig-

TABLE 2

ANALYSIS OF VARIANCE OF LOG TRIALS TO LEARN SECOND CONCEPT

Source	*df*	*F*
Reversal-nonreversal (R-NR)	1	41.32**
Consistent-inconsistent (C-IC)	1	17.34**
Number-position (N-P)	1	0.95
C-IC × R-NR	1	7.16*
N-P × R-NR	1	0.36
N-P × C-IC	1	1.27
N-P × R-NR × C-IC	1	1.53
Within groups (*MS*)	56	(0.187)

* $P < .01$.
** $P < .001$.

nificant interaction ($P < .01$) between the reversal-nonreversal and consistent-inconsistent groups, appears to be due to the much slower learning of the two nonreversal-inconsistent groups, as can be seen in Table 1.

Further breakdown of the reversal-nonreversal comparison was done by means of individual t tests. The transformed scores were used and the mean square within groups, obtained from the analysis of variance, was used as the basis for the error term. Two-tailed t tests showed that both kinds of reversal groups learned significantly faster than the comparable nonreversal groups. The reversal groups which shifted concepts within the same dimension of the stimuli, without making a literal reversal of their previous responses (nonliteral reversals), learned significantly faster than the comparable nonreversal groups ($t = 2.65$, $df = 56$, $P < .02$). The reversal groups which both shifted concepts within the same dimension of the stimuli and also made responses which were literal reversals of their previous responses, learned significantly faster than the comparable nonreversal groups ($t = 6.43$, $df = 56$, $P < .001$). The two types of reversal groups did not differ significantly from each other ($t = 1.05$, $df = 56$, $P > .05$). There was a significant difference between the two kinds of nonreversal groups ($t = 4.83$, $df = 56$, $P < .001$).

Discussion

The quicker learning of reversal shifts as opposed to nonreversal shifts was again found in this experiment. It should be remembered that in the present experiment the label, reversal group, was extended to all groups that learned a second concept which required discrimination according to a dimension of the stimuli that was previously relevant during Task 1 learning. It was found, in all cases, that a second sorting task which was based on the same dimension of the stimuli as the first task was learned quicker than comparable concepts which were not based on the same dimension of the stimuli. This occurred even when the reversal task did not require an exact literal reversal of previous sorting responses. Thus, the data support the previous analysis of Buss (1956), Harrow and Friedman (1958), and Kendler and D'Amato (1955), who hypothesized that a reversal shift is learned quickly because it has the advantage of using a dimension of the stimuli which was previously relevant. In a similar manner the results fit in with a mediational S-R approach.

The data also indicate that concepts requiring consistent responses are learned more quickly than concepts requiring inconsistent responses. These results are not surprising. It seems probable that *Ss* have frequently made other similar consistent responses before, in their daily lives, and that in the present experiment they quickly generalized to this particular situation.

The significant interaction between reversal-nonreversal and consistent-inconsistent groups suggests that the relative difficulty of reversal as opposed to nonreversal shifts is affected by whether the concepts used require consistent or inconsistent responses. Due to this, when both consistent and inconsistent responses are used in experiments of this type they should be controlled systematically, whether the interest is in the consistent and inconsistent responses (which in themselves have wide general applicability) or whether they are just used incidentally. In the present experiment it seems appropriate to analyze the significant interaction term with respect to the differences ($P < .001$) between the consistent and inconsistent nonreversal groups. Although the literal reversals involved inconsistent responses and the nonliteral reversals did not, the significant interaction appears to reflect the significant differences between the two types of nonreversal groups. Fitting this interpretation, the means of the two kinds of reversal groups did not differ

significantly from each other. It should also be noted, concerning inconsistent responses, that the groups required to make inconsistent responses were more sensitive to the experimental conditions. Thus, in similar card-sorting experiments it may be advisable, when practical, to use groups making inconsistent responses, due to their greater sensitivity.

Summary

The present experiment tested the notion that reversal shifts are learned more quickly than nonreversal shifts because they involve responding to a dimension of the stimuli which was used previously. This was accomplished in a four-category card-sorting situation in which both number and position concepts were used. Some concepts required consistent responses and other concepts required inconsistent responses. Sixty-four *Ss* learned two successive card-sorting tasks. During Task 1 all concepts required inconsistent responses. During Task 2 half of the *Ss* learned reversal tasks and half learned nonreversal tasks. Also, half the concepts used required consistent responses and half required inconsistent responses.

The results indicated that: (*a*) All types of reversal tasks (tasks requiring the use of a diminsion of the stimuli which was previously relevant) were learned in fewer trials than comparable nonreversal tasks (tasks requiring attention to a different dimension of the stimuli). Thus the previously reported hypotheses were supported. (*b*) Concepts requiring consistent responses were learned in fewer trials than concepts requiring inconsistent responses. (*c*) The relative difficulty of reversal shifts as opposed to nonreversal shifts is affected by whether consistent or inconsistent responses are used, i.e., there was a significant Reversal-Nonreversal × Consistent-Inconsistent interaction.

References

Berg, E. A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39,** 15–22.

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45,** 75–81.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52,** 162–166.

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

Gormezano, I., & Grant, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55,** 621–627.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55,** 592–598.

Kendler, H. H., & D'Amato, M. J. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49,** 165–174.

Kendler, H. H., & Mayzner, M. S., Jr. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51,** 244–248.

Kendler, T. S., & Kendler, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58,** 56–60.

(Received September 25, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 482-488

THE SERIAL POSITION EFFECT OF FREE RECALL[1]

BENNET B. MURDOCK, JR.

University of Vermont

Recently Murdock (1960) has shown that in free recall R_1, the total number of words recalled after one presentation, is a linear function of t, total presentation time. Nothing was said about the serial position effect, though this is a well-known phenomenon of free recall (e.g., Deese & Kaufman, 1957). However, given that there is a serial position effect, the simple linear relationship between R_1 and t is rather surprising.

In the customary serial position curve of free recall, probability of recall is plotted as a function of serial position. This means, then, that the area under the serial position curve is equal to R_1, the number of words recalled after one presentation. If R_1 is a linear function of t then it must follow that the area under the serial position curve is also a linear function of t. However, it is not immediately apparent how the serial position curve varies with t in such a way as to maintain this simple linear relationship.

The present experiment was designed as an attempt to determine how the serial position curve varied with list length and presentation rate while still maintaining this linear relationship. Unfortunately, at the end of the experiment it was still not clear how this relationship came about or, for that matter, whether the relationship was even linear after all. The basic reason for this failure was that the trends which did show up were not consistent enough to justify any clear-cut conclusions. However, a rather definite picture of the serial position curve itself did emerge from the data. Therefore, the present article will be restricted to a quantitative description and attempted explanation of the serial position curve of free recall.

[1] This study was supported by a research grant, M-3330, from the National Institutes of Health. The author would like to thank Ellen Lissner, Cynthia Marvin, and Frank Warhurst for analyzing the serial position data.

PROCEDURE

Six groups each had a different combination of list length and presentation rate. These six combinations were 10-2, 20-1, 15-2, 30-1, 20-2, and 40-1; the first number indicates list length and the second number indicates presentation time (in sec.) per item. Thus, 10-2 means a list of 10 words presented at a rate of 2 sec/item. Notice that the first two, middle two, and last two groups were matched for t, total presentation time (20, 30, and 40 sec., respectively).

For each group there were 80 different lists. The lists were constructed by randomly selecting words from the (approximately) 4000 most common English words (Thorndike-Lorge, 1944, G count of 20 and up), except that homonyms, contractions, and archaic words were excluded.

Group testing was used. Lists were read to *Ss* either at every beat (presentation rate of 1 sec/item) or at every other beat (presentation rate of 2 sec/item) of an electric metronome set at a rate of 60 beats/min.

After each list there was a recall period of 1.5 min. The *Ss* wrote down as many words as they could remember in any order that they wished. Each recall period was terminated by a verbal "Ready" signal which preceded the start of the next list by 5–10 sec. All groups were given 20 lists per session and four sessions; successive sessions were spaced 2–7 days apart. Nothing was said about rehearsing while the lists were being presented.

In all there were 103 *Ss*, students of both sexes from the introductory psychology course who were fulfilling a course requirement. Exact *Ns* by group are shown in Table 1.

Results

The data were first analyzed to determine if practice effects occurred over the four sessions. Analyses of variance showed that there was a significant ($P < .01$) improvement over the four sessions for Groups 10-2, 15-2, and 20-2; whereas the effect was significant at only the .05 level for Group 30-1 and was not significant ($P > .05$) for Groups 20-1 and 40-1. However, the largest difference obtained between the best and the worst session for any one group was 1.13 words, and all other intersession differences were less than 1.0 words. Therefore, when this practice effect is divided into four sessions and anywhere from 10 to 40 serial positions its effect on the serial position curves was negligible.

Table 1 shows the means and *SD*s of the number of words recalled per list (R_1). Each mean is based on 80 lists per *S* and from 15 to 19 *S*s per group. As predicted, groups with the same total presentation time did not differ significantly in mean number of words recalled. That is, no significant differences were found between Groups 10-2 and 20-1 ($t = 1.39$), between Groups 15-2 and 30-1 ($t = 1.00$), or Groups 20-2 and 40-1 ($t = 0.48$).

TABLE 1

Mean Number of Words Recalled

Group	*N*	Mean	*SD*
10-2	18	6.39	0.76
20-1	16	6.87	1.16
15-2	19	8.25	1.40
30-1	19	8.82	1.98
20-2	15	8.53	2.08
40-1	16	8.24	1.08

The serial position curves are shown in Fig. 1. Probability of recall is plotted as a function of serial position. For greater generality, we would also like to use the data from studies by Murdock and Babick (1961) and Deese and Kaufman (1957). In the Murdock-Babick study there were 18 *S*s each tested on 80 different 25-1 lists. In the Deese-Kaufman study there were two groups of 16 *S*s each; one group was tested on 10 different 10-1 lists and the other group was tested on 10 different 32-1 lists. The

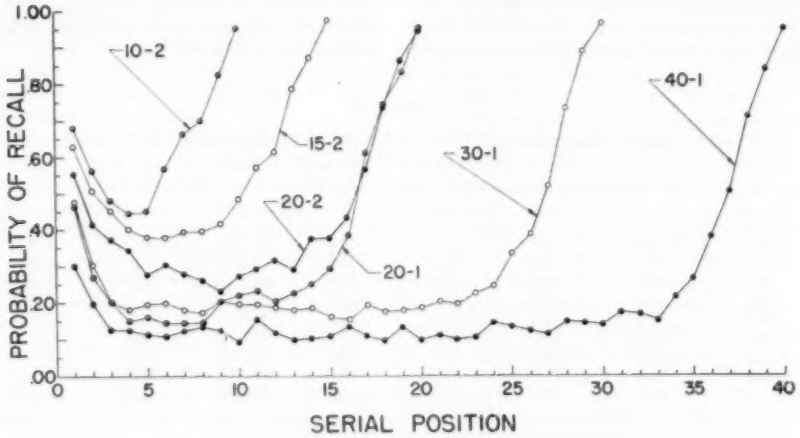

Fig. 1. Serial position curves for the six groups.

serial position data were presented in the original article as Fig. 1 (p. 182) and we read the points from the two curves as accurately as possible. These three serial position curves are shown here as Fig. 2.

We have, then, nine different serial position curves. In general, the curves seem to share certain general characteristics: a marked recency effect, a flat middle section, and a primary effect which is more precipitous though smaller in magnitude than the recency effect. The presence of a flat middle section, or asymptote, is clearest in the 40-1 list (Fig. 1), but becomes less and less obvious as list length decreases. Actually, in the two 10-word lists the primacy and recency curves may have intersected each other before an asymptote has been reached.

More specifically, the recency effect can adequately be described by the Gompertz double-exponential function. As given by Lewis (1960, p. 81) the equation is $y = vg^{h^x}$. Probability of nonrecall (y) was plotted as a function of list length minus serial position (x). Thus, the last word in a list would have an x value of 0, the next to last word an x value of 1, etc. Both v and g were fractional and positive. The asymptote v was determined from the mean recall probabilities averaged over the flat part on each serial position curve. The constants g and h were obtained by a least squares method described by Lewis (1960, pp. 82–88) using the last eight points of each serial position curve (except of the two 10-word lists where only the last four or five points could be used).

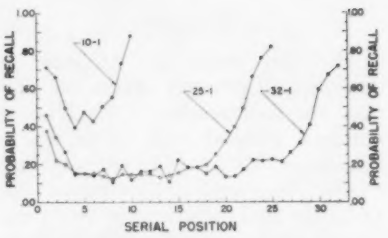

FIG. 2. Serial position curves for 10-1 and 32-1 lists (Deese & Kaufman, 1957) and 25-1 lists (Murdock & Babick, 1961).

The evidence for this conclusion is shown in Table 2 under the r^2 column. In all cases the Gompertz equation accounted for more than 95% of the variance, and the mean coefficient of determination (r^2) was 97.79%.

Since in all nine cases $g < 1/e$ the recency effect is consistently an **S** shaped curve. This characteristic can be seen in the serial position curves of Fig. 1 and 2. Starting from the last serial position, each curve is initially positively decelerated and then soon becomes negatively decelerated.

TABLE 2

VALUES FOR GOMPERTZ DOUBLE-EXPONENTIAL FUNCTION TO FIT SERIAL POSITION CURVES IN FREE RECALL

Group	v	g	y_0	h	r^2	x_i	$x_{.95}$
10-2	.548	.100	.055	.574	97.3%	1.5	6.9
20-1	.852	.050	.043	.596	98.8%	2.1	7.9
15-2	.622	.026	.016	.518	97.8%	2.0	6.5
30-1	.814	.032	.026	.546	99.3%	2.0	7.0
20-2	.730	.048	.035	.552	95.9%	1.9	6.9
40-1	.885	.036	.032	.557	98.7%	2.0	7.1
25-1[a]	.851	.134	.114	.634	98.3%	1.5	8.1
10-1[b]	.566	.206	.117	.431	98.3%	0.5	4.1
32-1[b]	.840	.270	.227	.644	95.7%	0.6	7.4

[a] From Murdock and Babick (1961).
[b] From Deese and Kaufman (1957).

The y_0 column gives the value of y when $x = 0$. If y_0 is subtracted from 1.00 this gives the probability that the word in the last serial position will be correctly recalled. The results for the six groups of the present experiment were very similar to each other, and an analysis of variance of the number correctly recalled showed that the groups did not differ significantly ($F = 1.61$, $df = 5/97$, $P > .05$). The recall probabilities were rather high but they were not 1.00 (and had they been the Gompertz would not be applicable); the corresponding recall probabilities for the Murdock-Babick and Deese-Kaufman data were clearly lower.

The inflection point occurs between the second and third words from the end of the list and appears to be essentially independent of list length and presentation rate. The evidence for this conclusion is given under the x_i column of Table 2, where $x_i = -\ln(-\ln g)/\ln h$ ($\ln$ is log base e). That is, x_i is the inflection point, that x value at which the deceleration changes from positive to negative. The x_i values range from 0.5 words to 2.1 words with a mean of 1.57 words. Since the last word in any list has an x value of 0, a mean of 1.57 words places the inflection point midway between the second and third words from the end of the list.

Actually, both Deese-Kaufman curves appear to have inflection points nearer the end of the list than any of the other curves. Otherwise, however, the inflection points cluster rather closely in the range of 1.5–2.1 words.

The recency effect extends over the last eight serial positions and appears to be essentially independent of list length and presentation rate. The evidence for this conclusion is given under the $x_{.95}$ column, where $x_{.95}$ is that value of x at which the curve is 95% down. That is, at this point forgetting is 95% of the asymptotic value. The 95% level serves as a convenient criterion to mark the end of the recency effect.

The mean of the $x_{.95}$ column is 6.88 words or, rounded off to the nearest whole number, 7 words. Except for the Deese-Kaufman 10-1 list all the values seem to be very close to 7 words. Since the x value is 7 words, the recency effect extends over the last eight serial positions.

Another way of indicating the similarity among different lists is by the h column of Table 2. In the Gompertz the constant h determines the rate of change. Since the values of h are all rather similar this indicates that all curves have a similar rate of change, and if they have a similar rate of change all curves should level out at about the same x value if the numerical values of g do not differ too greatly.

The primacy effect appears to extend over the first three or four serial positions. This can be seen in the serial position curves of Fig. 1 and 2, as all of the curves seem to level out at about the third or fourth serial position. The primacy effect is so short-lived that the curve is difficult to describe mathematically. Actually, it may well be exponential. Semilog plots of the first three or four points of the nine curves (using $1.00 - v$ as the asymptote for each curve) gave reasonable approximations to straight lines and the slopes were rather similar to one another. A group curve based on the mean $(y-c)$ values of the individual curves was an excellent fit; the rate constant was 0.77 and the intercept was .27 (see Murdock & Cook, 1960). However, the fact that this group curve was based on only three points should make one hesitant about placing too much confidence in it.

Finally, the primacy and recency effects are spanned by a horizontal asymptote. The asymptote is considered to extend from Serial Position 5 up to the last eight serial positions. That is, in a 20-word list the asymptote would extend from Serial Position 5 through Serial Position 12, in a 30-word list from Serial Position 5 through Serial Position 22, etc. That the asymptote is essentially horizontal is suggested by the middle parts of the serial position curves of Fig. 1 and 2.

A close examination of the serial position curves suggests that the trend line may have a small positive

TABLE 3

PREDICTED AND OBTAINED INCREMENTS FOR ASYMPTOTE

Group	ΔX	Pred.	Obt.	Diff.
20-1	8	.039	.100	.061
30-1	18	.022	−.004	−.026
20-2	8	.019	.015	−.004
40-1	28	.027	.038	.011
25-1	13	.044	.036	−.008
32-1	20	.032	.046	.014

slope rather than a zero slope. However, this positive slope could be due to the fact that the recency effect is only 95% down; i.e. 5% of the effect remains to exert an effect on the (allegedly) horizontal asymptote. The proper test of this conclusion, then, is to determine whether the obtained increment (if any) is greater than the increment attributable to the 5% remaining from the recency effect.

The following analysis deals only with lists of 20 words or more; the 10 and 15 word lists could not be used because there were too few points. For each of the six lists the obtained increment was found by fitting a least squares regression line to the asymptote, determining its slope, then multiplying the slope by Δx where Δx is the difference between Serial Position 5 and the seventh-from-last serial position ($\Delta x = 8$ for the 20-word list, $\Delta x = 13$ for the 25-word list, etc.). The expected increment was found by obtaining the predicted y value from the Gompertz equation for the two values of x (Serial Position 5 and seventh-from-last serial position), then subtracting. For each list the constants shown in Table 2 were used. The predicted and obtained increments are shown in Table 3; the difference between the predicted and obtained increments was not statistically significant ($t = 0.72, df = 5$). Thus, the asymptote does appear to be horizontal, and the slight positive slope to the curve is no greater than would be expected from the tail end of the recency effect.

DISCUSSION

We have presented data to show that the serial position curve of free recall is characterized by a rather steep (possibly exponential) primacy effect, an **S** shaped recency effect, and a horizontal asymptote extending between the primacy and recency effect. An idealized curve for a 24-word list is shown in Fig. 3. Its equation is

$$p = 1.00 + .27\, e^{-.77(x-1)} - .772\,(.042)^{.555(L-x)}$$

where L is list length and x is Serial Position 1, 2, 3, . . . L. The constants for the primacy effect were those of the group curve discussed above while the constants for the asymptote and the recency effect were the mean values of the constants given in Table 2 for the six lists of the present experiment.

The curve of Fig. 3 is an empirical curve, not a rational curve. It is an attempt to describe the serial position effect of free recall quantitatively, not explain it. Not only does this empirical curve represent the nine curves of Fig. 1 and 2 quite well, but also it is consistent with several other sets of data. For one, it agrees with serial position curves for 20-1 lists reported by Deese (1957,

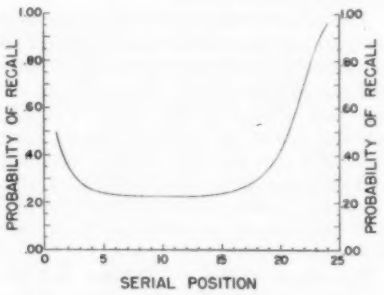

FIG. 3. Idealized serial position curve for 24-word list.

Fig. 1, p. 580). For another, it agrees well with some unpublished curves culled from several experiments recently reported by Murdock (1960). Finally, the exact same trends are present in some memory-span data reported by Waugh (1960, Fig. 3, p. 75).

However, the empirical curve of Fig. 3 is not in agreement with results reported by Bousfield, Whitmarsh, and Esterson (1958). These authors used 5-, 10-, 20-, and 40-word lists all presented at a rate of 2.5 sec/word, and consistently found the primacy effect more marked than the recency effect. Both Bousfield et al. (1958, pp. 260–261) and Deese (1957, pp. 581–582) suggest that the relatively slow presentation rate may have encouraged rehearsal and thus led to the greater primacy effect. To investigate this possibility, we conducted an additional experiment with 35 *Ss* using 10 20-2.5 lists. The 20-word length was selected because the curves of Bousfield et al. (1958, Fig. 1, p. 258) seemed to show the most pronounced primacy effect for this length list. As Bousfield et al. (1958) apparently used a somewhat longer recall period we used a 4-min. recall period in this additional experiment; otherwise the procedure was identical with that of the other experiments reported here.

The results of the experiment are shown in Fig. 4. As can be seen, in general the results are quite consistent with the empirical curve of Fig. 1, and in particular the recency effect is more pronounced than the primacy effect. This experiment clearly shows that the results of Bousfield et al. (1958) are not due to the slower presentation rate per se.

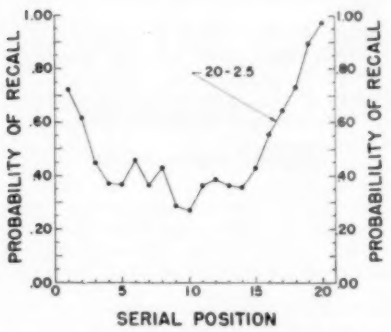

FIG. 4. Serial position curve for 20-2.5 lists.

Why did Bousfield et al. (1958) find primacy more pronounced than recency? One possibility is their instructions. Twice in their instructions they told *Ss* that the words were to be recalled, ". . . in the order in which they occur in your memory." The stress on order may have given *Ss* a set to recall the words in the order presented, and Deese (1957) has shown that instructions to *Ss* are an important variable in determining the shape of the curve. A second possibility is the design used. Bousfield et al. (1958) used a counterbalanced design such that each *S* had only one list at each length. Thus, in effect each list was (to *S*) of unknown length, and this fact may have encouraged rehearsal in the order of presentation.

In any event, under the conditions of the present experiment there seems little doubt that the serial position effect of free recall is essentially as depicted in Fig. 3. Of course, as Deese (1957, p. 581) has noted, the serial position curve is sensitive to the introduction of experimental variables. However, it has been found that more items are recalled with free recall than with ordered recall (Deese, 1957; Waugh, 1961), so evidently free recall is the preferred, perhaps even the more basic, method of recalling a list of unrelated words.

Finally, why does the serial position curve of free recall take the shape it does? One possible explanation is in terms of short-term proactive and retroactive inhibition. That is, each word in a list is both preceded by anywhere from 0 to $(L - 1)$ other words and followed by anywhere from $(L - 1)$ to 0 other words. Up to a point, the more preceding words the more short-term PI and the more succeeding words the more short-term RI. The PI and RI effects presumably summate to determine the total inhibitory effects.

If this explanation is correct, recent

studies of the short-term retention of individual items should provide an indication of the course of PI and RI to be expected. It has been found that, in short-term memory, PI effects appear to be greatest after about three prior words (Murdock, 1961). This agrees well with the finding that the primacy effect levels out after the first three or four serial positions. In short-term memory RI effects appear to approach an asymptotic value greater than zero (Murdock, 1961; Peterson & Peterson, 1959). This agrees well with the finding of a horizontal asymptote in the serial position curve. Finally, an examination of the RI curve of short-term memory even suggests an **S** shaped curve (see proportion of correct recalls over different retention intervals, Tables 1 and 3, Murdock, 1961, pp. 619–620). This agrees well with the Gompertz recency effect suggested here. Thus, it would appear that all the main characteristics of the idealized serial position curve shown in Fig. 3 are compatible with the results obtained from the short-term retention of individual items, and these findings lend support to the idea that the serial position curve of free recall is essentially a manifestation of short-term PI and RI effects.

Summary

This experiment was a study of the serial position effect of free recall. Curves were obtained for 10-2, 20-1, 15-2, 30-1, 20-2, and 40-1 lists, where the first number indicates list length and the second number indicates presentation time per word. On the basis of the available evidence it was concluded that, under the conditions of the present experiment, the serial position curve is characterized by a steep, possible exponential, primacy effect extending over the first three or four words in the list, an **S** shaped recency effect extending over the last eight words in the list, and a horizontal asymptote spanning the primacy and recency effect. Finally, it was suggested that the shape of the curve may well result from proactive and retroactive inhibition effects occurring within the list itself.

References

Bousfield, W. A., Whitmarsh, G. A., & Esterson, J. Serial position effects and the "Marbe effect" in the free recall of meaningful words. *J. gen. Psychol.*, 1958, **59**, 255–262.

Deese, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

Deese, J., & Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

Lewis, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

Murdock, B. B., Jr. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

Murdock, B. B., Jr. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625.

Murdock, B. B., Jr., & Babick, A. J. The effect of repetition on the retention of individual words. *Amer. J. Psychol.*, 1961, **74**, 596–601.

Murdock, B. B., Jr., & Cook, C. D. On fitting the exponential. *Psychol. Rep.*, 1960, **6**, 63–69.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Waugh, N. C. Serial position and the memory-span. *Amer. J. Psychol.*, 1960, **73**, 68–79.

Waugh, N. C. Free versus serial recall. *J. exp. Psychol.*, 1961, **62**, 496–502.

(Received October 6, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 489–494

THE SCALING OF SUBJECTIVE ROUGHNESS AND SMOOTHNESS[1]

S. S. STEVENS AND JUDITH RICH HARRIS

Harvard University

These experiments began as an attempt to apply the method of magnitude estimation to a continuum for which the stimulus seemed to have no metric scale, only an ordinal scale of grades of sandpaper, or emery cloth. Unexpected discoveries led on to more engaging inquiries. At the outset, both a ratio scale and a category scale of apparent tactual roughness were determined with 12 grades (grits) of emery cloth. The relation between the ratio scale and the category scale was typical of the relation found on prothetic continua. (For these first results, see Stevens, 1961a; for an earlier related study, see Dudek & Baker, 1956.)

The next study was an exercise in which two students, C. S. Harris and J. P. McMahon, asked 12 *Os* to judge the smoothness of the stimuli instead of the roughness. The ratio scale of smoothness approximated the inverse, or reciprocal, of the ratio scale found for roughness, and the category scale of smoothness was the reverse, or the complement, of the category scale of roughness. These results resemble Torgerson's (1960) findings when he scaled both the apparent lightness and the apparent darkness of gray papers.

In terms of a linear scale of apparent roughness, it turned out that the stimuli used were bunched rather tightly at the low (smooth) end of the continuum, so much so that the two category scales were almost logarithmic functions of the respective ratio scales. Other studies of category scaling suggest that the nearly logarithmic form found here is an accident of the stimulus spacing, and that if an iterative procedure were used to arrive at a "pure" category scale the curve for roughness would be less curved than a logarithmic function (Stevens & Galanter, 1957).

Since the available number of emery cloth grits is limited, it is difficult to determine a pure category scale, but when a sample of grits more uniformly spaced in subjective roughness was used, the form of the category scale changed as predicted: it became much less curved than a logarithmic function when plotted against the ratio scale of subjective roughness. Ten *Os* judged grits 320, 120, 80, 50, 40, 30, and 24 twice each on a seven-point scale. The average judgments were 1.17, 2.17, 3.04, 4.12, 5.17, 6.08, and 6.62. These values determine a line that is straighter than a logarithmic function—a line that is not far from the pure form of the category scale, as evidenced by the tendency of *Os* to use each category number approximately equally often (Stevens & Galanter, 1957). It appears, therefore, that roughness behaves like a prothetic continuum, and that the pure category scale is not a logarithmic function of the magnitude scale (Eisler, 1962).

From these preliminary studies a surprising fact emerged: magnitude estimations of roughness, and of smoothness, turned out to give fairly straight lines when plotted in log-log coordinates against the grit number of the emery cloths. Another instance, it seems, of the psychophysical power law. (Grit number refers to the number of openings per inch in the screen employed to sift the abrasive particles.) If apparent roughness and its reciprocal, apparent smoothness, are power functions of particle size, it becomes a challenging task to determine more accurately the exponents involved. In the preliminary experiments, each involving 12 *Os*, the approximate exponents were -1.5 for roughness and $+1.2$ for smoothness when measured against grit number. The next problem was to determine

[1] This research was supported by Grant G-10716 from the National Science Foundation (Report PAR-266).

which of these two exponents, if either, is more nearly correct in absolute value.

Apparatus and Procedure

The stimuli were the twelve grits, 320, 240, 220, 180, 120, 100, 80, 60, 50, 40, 30, and 24. It was assumed that these grits, Tri-M-ite brand, met the published standards for abrasives (Horton, 1957) which allow the grain size to vary with a standard deviation of approximately 20% around the nominal size. Two different sets of cloths were used. To a sufficient approximation for the present purpose, the grit number can be regarded as proportional to the reciprocal of the grain diameter.

The stimuli were presented one at a time to *O*, who placed his hand through a cloth-screened opening. He stroked the emery cloth twice with his index and middle fingers. Two samples of each of the 12 grits were presented in a different irregular order to each *O*.

Two experiments used the method of magnitude estimation, and one used cross-modality matching, as follows.

Experiment 1 (with assigned modulus).—Grit 100 was presented first and *O* was told to call it 10. Of the 20 *O*s, 10 first judged roughness and on a later date judged smoothness. The other 10 reversed the order.

Written instructions were given to each *O*. When roughness was being judged, the instructions were:

> I am going to present a series of surfaces that vary in roughness. Your task is to tell me how rough they feel by assigning numbers to them. The first will be the standard roughness, which we will call 10. Your task is to assign numbers proportional to your subjective impression. Use whatever numbers seem appropriate—fractions, decimals, or whole numbers. For example, if a surface feels 3 times as rough as the standard, say 30; if it feels ½ as rough, say 2, etc. Try not to worry about being consistent; try to give the appropriate number to each surface regardless of what you might have called some previous surface. In feeling the surfaces, draw your index and middle fingers twice across each surface as it is presented.

When the task was to judge smoothness, the words "smooth" and "smoothness" were substituted for "rough" and "roughness" in the preceding instructions.

Experiment 2 (no assigned modulus).—The first stimulus presented was different for each *O*, and, instead of there being a modulus called 10, the instruction was to call the first stimulus "any number you think appropriate." Roughness and smoothness were judged on different days by each of 10 *O*s.

Experiment 3 (cross-modality matching).—The *O* adjusted the intensity of a band of noise (500 to 5000 cps) until its subjective magnitude appeared to match the subjective magnitude of the roughness (or smoothness) of each grade of emery cloth. Two matches were made to each of the 12 cloths in an irregular order by 10 *O*s. One group of 10 *O*s matched loudness to roughness; another group of 10 *O*s matched loudness to smoothness.

The loudness was controlled by a "sone potentiometer" (two 2000-ohm potentiometers, ganged and cascaded). An additional attenuator in series enabled *E* to keep *O*'s adjustments of the sone potentiometer more or less centered in the usable range. The voltage across the PDR-8 earphones was measured with a vacuum-tube voltmeter.

Results

The geometric means of the results of the 20 *O*s who judged roughness and smoothness in terms of an assigned modulus (Grit 100 called 10) are shown in Fig. 1. The slopes of the two straight lines are equal but of opposite sign (+1.4 and −1.4). This is what is called for by the reciprocal relation between roughness and smoothness. On the other hand, the points do not always fall close to the lines. In general, there is a tendency for both functions to be slightly concave downward.

It is also of interest that the standard called 10 at the beginning of each run was judged less rough when it was presented again as a stimulus to be judged (geometric mean = 6.43). When the task was to judge smoothness, the standard called 10 was later judged more smooth (geometric mean = 13.03). A similar adaptation—if that is what it should be called—was also noted in the preliminary experiments.

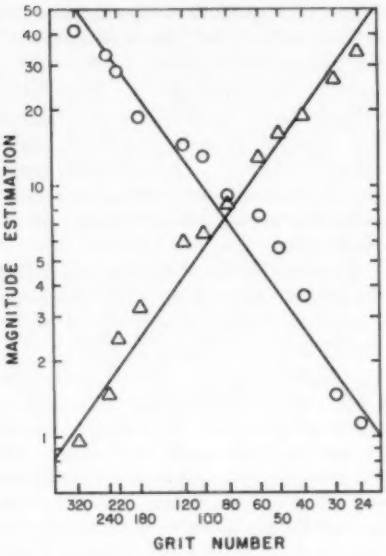

FIG. 1. The geometric means of the estimations of roughness (triangles) and smoothness (circles) are plotted against grit number in log-log coordinates. (Each point is based on 40 judgments—20 *Os*.)

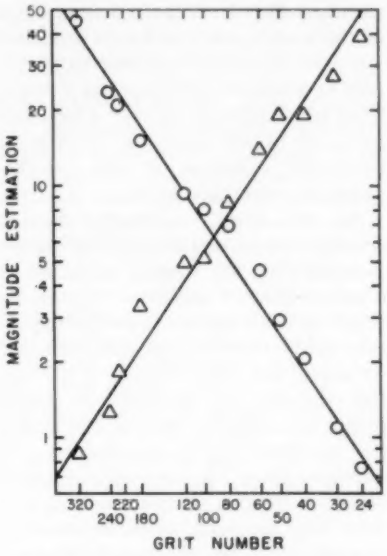

FIG. 2. Similar to Fig. 1, except that in these experiments on roughness (triangles) and smoothness (circles) no standard stimulus was designated by *E*. (Each point is based on 20 judgments—10 *Os*.)

Experiment 2, with no assigned modulus, gave the results shown in Fig. 2. The points (geometric means) lie closer to the straight lines and the slopes are ± 1.5, a slightly greater absolute value than in Exp. 1. Allowing each *O* to choose his own modulus appears to have produced a better result. This added freedom has led *Os* to give superior results on other occasions when no standard was used (Stevens, 1956). Other things being equal, it is better in experiments with magnitude estimation to dispense with an assigned modulus and to average by taking the geometric means of the judgments. No prior processing of the data is necessary with geometric averaging.

The results of matching loudness to roughness and to smoothness are

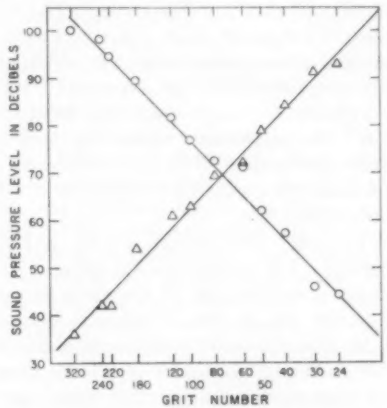

FIG. 3. Sound pressure levels in decibels re 1 microbar produced in the earphones when *Os* matched the loudness of a noise to the roughness (triangles) and the smoothness (circles) of the 12 grades of emery cloth. (Each point is the decibel average of 20 judgments—10 *Os*.)

shown in Fig. 3. The decibel averages of the sound pressure levels produced when *O*s matched loudness to roughness and smoothness produce straight lines when plotted against a logarithmic scale of grit numbers. This is the predicted outcome if the psychophysical power law holds.

As has been repeatedly demonstrated (Stevens, 1959, 1961b) the exponents (log-log slopes) obtained in cross-modality matches should be equal to the ratio between the exponents of the two modalities determined by magnitude estimation. If we take the exponent for loudness vs. sound pressure to be 0.6 (Stevens, 1955), and the exponent for roughness vs. grit diameter to be 1.5, the predicted slope of the matching function in Fig. 3 is 2.5. The measured slope is 2.6. For smoothness, the measured slope is -2.6, as expected. In view of the sources of error and uncertainty in these experiments, it is reassuring that the measured and the predicted exponents in cross-modality matching agree within 4%. The cross-modality matches suggest that the roughness-smoothness exponents may be slightly greater than 1.5. If we could vary roughness to match loudness, presumably the exponent would be larger still (Stevens, 1959) but this complementary experiment would be difficult to execute.

In order to provide a unit and a formula for subjective roughness, it appears reasonable for the time being to take 1.5 as the exponent and to take the apparent roughness of Grit 320 as the subjective unit. If a name for this unit is desired, the term *ruk*, derived from a root cognate of rough, is suggested. Accordingly, the equation for subjective roughness R in ruks, as a function of grit number G becomes:

$$R = 5724G^{-1.5}$$

In terms of the average diameter in millimeters of the abrasive particles, the exponent would be $+1.5$, and the constant would be 106.5. This latter value is based on the sieve openings in the United States standard series (Horton, 1957). The equation omits the "threshold" constant, because the available stimuli did not permit its evaluation.

The main uncertainties regarding the size of the exponent for roughness stem from two causes, each of which would have an opposite effect. The available range of stimuli was relatively short, about 1.12 log units, a factor that would be expected to increase the measured exponent. On the other hand, the experiments involved the matching of numbers (or of loudnesses) to roughness (or smoothness), never the reverse. Experience has shown that this unbalanced design tends to decrease the measured exponent. Whether the two sources of presumed bias have equal as well as opposite effects cannot be told without further experimental analysis.

Discussion

Reciprocality.—It is clear from Fig. 1, 2, and 3 that values obtained in the judgment of smoothness are approximately the reciprocals of values for apparent roughness. It seems likely that any continuum can be judged with at least fair success in terms of its reciprocal, although only a few have been looked at from this point of view. In addition to the continuum roughness–smoothness, there are data on lightness–darkness (of surfaces) and loudness–softness (of noises). One can easily imagine judging the longness or shortness of lines, the strength or weakness of vibrations, the brightness or dimness of luminances, the heaviness or lightness of lifted weights, and so on. The inverting of a continuum is a kind of semantic matter: *O* is instructed differently, and he tries to

respond with numbers that are inversely proportional to those he would use for the continuum "right side up." At least he tries to report reciprocals provided he understands the semantic rule: a surface that is twice as rough is half as smooth.

Some continua, like roughness, are frequently referred to in terms of the inverse aspect, but for many continua *O* would probably find it a little strange to be told to judge the reciprocal. Even with smoothness, a few *O*s commented on the difficulty of the task. It seemed easier and more natural, for example, to match loudness to roughness than to match loudness to smoothness.

Even though people may become accustomed to judging in terms of the inverse aspect, there is a sense in which what may be called the degree of stimulation is basically different from the degree of its absence. When he judges the degree of absence of stimulation, *O* manages, with fair success, to report the reciprocal of his judgment of the strength of the sensory magnitude. A related type of report was tried out by E. C. Poulton who asked *O*s to estimate fractional loudness by varying only the denominator of their report (see Stevens, 1956). As the tones became fainter the numbers became larger in a manner that produced a nice reciprocal relation to the standard results of magnitude estimation. There are many ways of asking *O* to report the apparent magnitudes of a series of sensory excitations, but it seems hardly likely that all methods of report that are logically equivalent will give equally good results.

It is also important to note that some sensory attributes that may be thought of as opposites are not at all reciprocally related. Warm and cold are two striking examples (Stevens & Stevens, 1960). The exponent for warm is 1.6, for cold 1.0. These seem to be two different sense modalities, not two names for the same continuum, and the stimulus domains of the two continua do not overlap. Presumably experiments could be run in which *O* told how warm or how neutral the stimulus felt, and the results would follow the reciprocal relation. An analogous experiment could be run with *O* judging cold or neutral, again with the expectation of obtaining reciprocal functions, but the slopes (exponents) would differ greatly from the slopes obtained with the warm-neutral pair. One must distinguish, therefore, between word pairs that refer to two different continua and word pairs that refer reciprocally to a single continuum. A given temperature is usually *either* warm or cold, but a given surface is *both* rough and smooth.

The stimuli.—The stimuli used in these studies were those samples of emery cloth that happened to be available in the laboratory. They appeared to be in good order, and they seemed to have been a good set of stimuli, in the sense that they gave rise to orderly power functions when used in eight different experiments. It is noteworthy, however, that, as the experiments accumulated, the evidence became increasingly clear that there were minor peculiarities in the series of grits, or perhaps in the manner of their adhesion to the cloths. It is not usual to argue from the results of magnitude estimation to a nonuniformity in a manufacturing process, but the evidence for at least one such "defect" seems clear. The cloth with Grit 120 was consistently judged too rough relative to Grit 100. This fact is especially evident in Fig. 2 and 3 and was also clear in the preliminary experiments.

Under the microscope one can see that Grit 120 involves smaller abrasive grains than Grit 100, as indeed it should, but the two cloths differ in the degree to which the particles appear to be immersed in the adhesive. The finer particles (Grit 120) appear to sit higher on the cloth and to present more of their surface to view. To the touch, the two cloths (100 and 120) feel different in a way that can best be described as qualitative. The skin catches occasionally on the finer particles in a way that is not characteristic of the coarser particles. This catching, due presumably to a shallower immersion of the particles in the bonding adhesive, may account for the relatively higher numerical estima-

tions of the apparent roughness of the finer grit.

Whatever the explanation, it becomes clear that particle size is not the only variable in a bonded abrasive that can affect apparent roughness. Nevertheless, the samples used in these experiments seem to have been only minimally contaminated by other factors, for sensed roughness grows as a relatively clean power function of average particle diameter.

Summary

Preliminary experiments showed that *Os* can make consistent judgments of tactual roughness and smoothness. The stimuli were 12 grits of emery cloth. Magnitude estimations of roughness and smoothness produced straight lines when plotted (log-log) against grit number. The exponents of these power functions were determined in two experiments with magnitude estimation and one with cross-modality matching against loudness. All three experiments gave results that were power functions of grit number with exponents in the vicinity of −1.5 for roughness and +1.5 for smoothness. The cross-modality matches also confirmed the exponents determined by magnitude estimation.

References

Dudek, F. J., & Baker, K. E. The constant-sum method applied to scaling subjective dimensions. *Amer. J. Psychol.*, 1956, **69**, 616–624.

Eisler, H. Empirical test of a model relating magnitude and category scales. *Scand. J. Psychol.*, 1962, **3**, 88–96.

Horton, H. L. (Ed.). *Machinery's handbook.* (15th ed.) New York: Industrial Press, 1957. P. 1459.

Stevens, J. C., & Stevens, S. S. Warmth and cold: Dynamics of sensory intensity. *J. exp. Psychol.*, 1960, **60**, 183–192.

Stevens, S. S. The measurement of loudness. *J. Acoust. Soc. Amer.*, 1955, **27**, 815–829.

Stevens, S. S. The direct estimation of sensory magnitudes: Loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

Stevens, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, 1959, **57**, 201–209.

Stevens, S. S. To honor Fechner and repeal his law. *Science*, 1961, **133**, 80–86. (a)

Stevens, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 1–33. (b)

Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen and S. Messick (Eds.), *Psychological scaling.* New York: Wiley, 1960.

(Received October 21, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 495–504

AN EVALUATION OF THE ACTIVATIONIST HYPOTHESIS OF HUMAN VIGILANCE[1]

JACK A. ADAMS AND LAWRENCE R. BOULTER

University of Illinois

The arousal or activationist hypothesis of human attentiveness holds that the stimuli impinging on *S* from external sources, or acting from within him, determine alertness through the reticular activating system of the brain (Broadbent, 1958; Fiske, 1961; Frankmann & Adams, 1962; Lindsley, 1957; Malmo, 1959). Studies of human vigilance, or long-term attentiveness for occasional signals in a monitoring task, commonly show a decrement in signal detection as a function of observation time, and Scott (1957) interprets this phenomenon as adaptation to the nearly unchanging stimulation. Frankmann and Adams (1962) have pointed out that vigilance decrement is not as ubiquitous as is sometimes believed, and that investigators have failed to ask why decrement is more prominently associated with simple tasks and is absent or small in complex visual tasks with multiple stimulus sources. In terms of the activationist hypothesis, there appear to be special sources of stimulation in complex tasks.

Efforts to use explicitly the hypothesis in a careful predictive sense for vigilance findings are hampered by an absence of relationships between the locus, amount, and type of stimulation, and measures of molar behavior (Adams, Stenson, & Humes, 1961; Frankmann & Adams, 1962). However, with the hypothesis firmly rooted in studies of the reticular formation, there is a number of physiological findings that can be used to facilitate a search for the missing definitions and relationships at the molar level. The two experiments reported here reason from physiological findings and seek to identify sources of stimulation that can deter decrement in complex visual tasks. One derivation was that a special source of stimulation in complex tasks can be the proprioceptive stimulation associated with the head and eye movements of scanning the stimulus array and it is based on proprioception collaterals in the reticular formation (Rossi & Zanchetti, 1957; Samuels, 1959). When the stimulus sources have wide spatial separation, head movements would be the most prominent source of proprioceptive stimulation, but even when multiple stimulus sources are quite close together we can still expect proprioceptive stimulation from eye movements alone (Cooper, Daniel, & Whitteridge, 1955; Whitteridge, 1960). The changing retinal stimulation as the head and eyes scan the visual scene is an added source of stimulation.

Another derivation was that response complexity can produce internal stimulation that should decrease the amount of decrement, and its interaction with the proprioception and visual variables was included to illuminate the effects on vigilance of another source of stimulation in complex tasks. Physiologically, complex responses might be expected to

[1] This research was supported by Contract AF 19(604)-5705 monitored by the Operational Applications Laboratory, Deputy for Technology, Electronic Systems Division, Air Force Systems Command.

influence alertness through cortical outputs fed back to the reticular system (Adams et al., 1961; Lindsley, 1957; Rossi & Zanchetti, 1957; Samuels, 1959). Adams and Boulter (1960) and Adams, Stenson, and Humes (1961) have shown that a relatively complex decision response of four choices associated with each signal detection will decrease the amount of vigilance decrement.

Experiment I

Method

Apparatus.—Two units of the multiple source vigilance apparatus were used, with 2 *S*s tested at a time. The *S* sat at a desk and faced a semicircular track fixed at eye level to the back of the desk. On the track, and 30 in. from *S*'s eyes, were mounted four small ($1\frac{1}{2} \times 2$ in. front surface) digital display boxes that presented a two-digit number ($\frac{5}{16}$ in. high) which was the critical signal *S* had to detect. Six different numbers were used. The number was always bright enough for easy reading, but the brightness level was intentionally low so *S* had to orient to a stimulus source to see the signal easily. When the critical signal came on, *S* was instructed to report the event as fast as possible by pressing a detection button located 4 in. from his fingertips. A timer started at the onset of the signal and stopped when the detection button was touched, thus giving a response latency score. This simple mode was called *detection responding*. Under another experimental condition called *memory responding*, which is described in more detail below, *S* had a set of six memory buttons in addition to the detection button. These buttons were each labeled with one of the numbers and were arrayed in a semicircle around the detection button, and 4 in. from it. After *S* pressed the detection button as rapidly as possible, he further selected and pressed a memory button, also as rapidly as possible. The latency from the onset of the signal to pressing the memory button was recorded too. An *S* responded with his preferred hand, but he always kept both forearms in an armrest to standardize the position of the arm and hand. Latency measures such as these have been found to have wide variability when the starting position for the hand is not controlled. A microswitch under each armrest turned on a light at *E*'s panel whenever *S* removed an arm from the armrest and, when this occurred, *E* administered a brief reminder over an intercom system. An occasional reminder in the practice session, which preceded criterion sessions, was usually sufficient to standardize this procedure. All signals and intersignal intervals were programed and automatically read and timed by a digital tape reader that fed the two apparatus units simultaneously. Each *S* was in his own experimental room, which had normal illumination. The *E* and his console were located in another building.

Description of stimulus series.—Each session was 2.5 hr. long, and for scoring purposes was divided into five 30-min. trials. There were eight signals on a trial, and each signal lasted for 5 sec. The signals were the numbers 10, 20, 30, 40, 50, and 60, and each number occurred at least once on a trial. Two numbers appeared at each of the four stimulus sources on a trial. Intersignal intervals on each trial were .7, 1.2, 2.0, 3.0, 4.0, 5.0, 6.2, and 7.5 min. Both the assignment of intersignal intervals, and of signals to the four stimulus sources, were separately randomized for each trial. By repeated random sampling, a pool of 10 tapes was generated according to these rules.

Experimental procedures.—A major hypothesis was that sustained vigilance in complex visual tasks is a function of the proprioceptive and retinal stimulation arising from head and eye movements, and the manipulation of this class of stimulation was by spatial separation of the four stimulus sources. The assumption was that the wider the separation of sources, the greater the head and eye movements, and the greater the stimulation. Four amounts of spatial separation, specified in terms of the difference between the two outermost sources, were 18°, 36°, 72°, and 144°. The other two sources were spaced between the two peripheral ones to give equal separation between all four. Separations of 18° and 36° placed the four sources in a direct field of visual view, and essentially they could be scanned with eye movements alone. The 72° separation required some added head movements for comfortable scanning, and scanning the 144° separation was not possible without head movements.

A second major hypothesis was that response complexity is a source of internal stimulation for high alertness in complex tasks. Two levels of response complexity were used. One was detection responding, which was the simplest, that only required *S* to press the detection button when a signal was

detected. The other was memory responding, which had much more complexity, and it required *S* to actively employ his immediate memory throughout the 2.5 hr. The *S* always had to remember the last set of four numbers that appeared at the stimulus sources. As soon as he detected a new number at a source, his response was to press first the detection button and then press the memory button labeled with the number that had appeared at that same source the last time. This requirement for short-term memory was quite similar to running memory span (Pollack, Johnson, & Knaff, 1959), and it was hypothesized that the continuous memory requirement would introduce greater internal stimulation than detection responding.

Two groups of 15 *S*s each were used, with each *S* participating for five sessions. A mixed analysis of variance design was used (Lindquist, 1953, p. 292). Response complexity was a between-*S*s variable, with one group having detection responding throughout and another group having memory responding. Within-*S*s variables were five trials and four spatial separations, and were administered to all *S*s of both groups. Following a practice session with a randomly assigned spatial separation of sources, each *S* was given four criterion sessions where each session was a different one of the four degrees of spatial separation. So that *S* did not have two identical separation conditions in a row, the three separation conditions that were not used in the practice session were equally divided among the *S*s for the first criterion session. The separation condition for the practice session was assigned to the last criterion session, and the remaining two separations were randomly assigned to the second and third criterion sessions. To control for the possibility of some learning of the characteristics of a particular input tape over five sessions, a tape from the pool of 10 tapes was assigned randomly to an *S* for each session, with the restriction that no tape be used twice for any *S*. Each session was on a different day.

Subjects.—The 30 *S*s were male undergraduate students who were paid for their participation. They were randomly assigned to groups.

Results

The level of signal detection was high, as might be expected with a signal that persisted for 5 sec. Both detection and memory responding had a detection level of 98%.

Figure 1 shows the main findings. Each *S*'s score was the mean latency of his responses on a trial, and Fig. 1 presents the plot of group mean latencies as a function of trials. The latencies presented in Fig. 1 are for response to the detection button. Latencies increased as spatial separation increased, and memory responding had longer latencies than detection responding. The differences found between detection and memory responding were found to be primarily associated with pressing the detection button, indicating that all recall and decision delays about which of the six memory buttons to press took place before the main detection button was pressed. The difference between the latency to press the detection button and the latency to press the subsequent memory button is the time to move from the detection button to the memory button, and the means of these difference latencies

TABLE 1

ANALYSIS OF VARIANCE IN EXP. I

Source	*df*	*MS*	*F*[a]
Between *S*s			
Detection-Memory (C)	1	95.88	31.85*
$Error_b$	28	3.01	
Within *S*s			
Trials (A)	4	1.23	13.67*
Spatial separation (B)	3	20.08	59.06*
A × B	12	0.11	1.10
A × C	4	0.15	1.67
B × C	3	0.05	0.15
A × B × C	12	0.07	0.70
$Error_1(w)$	112	0.09	
$Error_2(w)$	84	0.34	
$Error_3(w)$	336	0.10	
Total	599		

[a] $Error_b$ used to test C effect. $Error_1(w)$ used to test A and A × C, $Error_2(w)$ used for B and B × C, and $Error_3(w)$ used for A × B and A × B × C.
* $P < .01$.

were found to be essentially constant for all conditions. These latter data have been omitted here.

Of primary interest was the influence of spatial separation and response complexity on decremental trends, and Fig. 1 shows decrement to be associated with all conditions. If our hypotheses about the effects of spatial separation and response

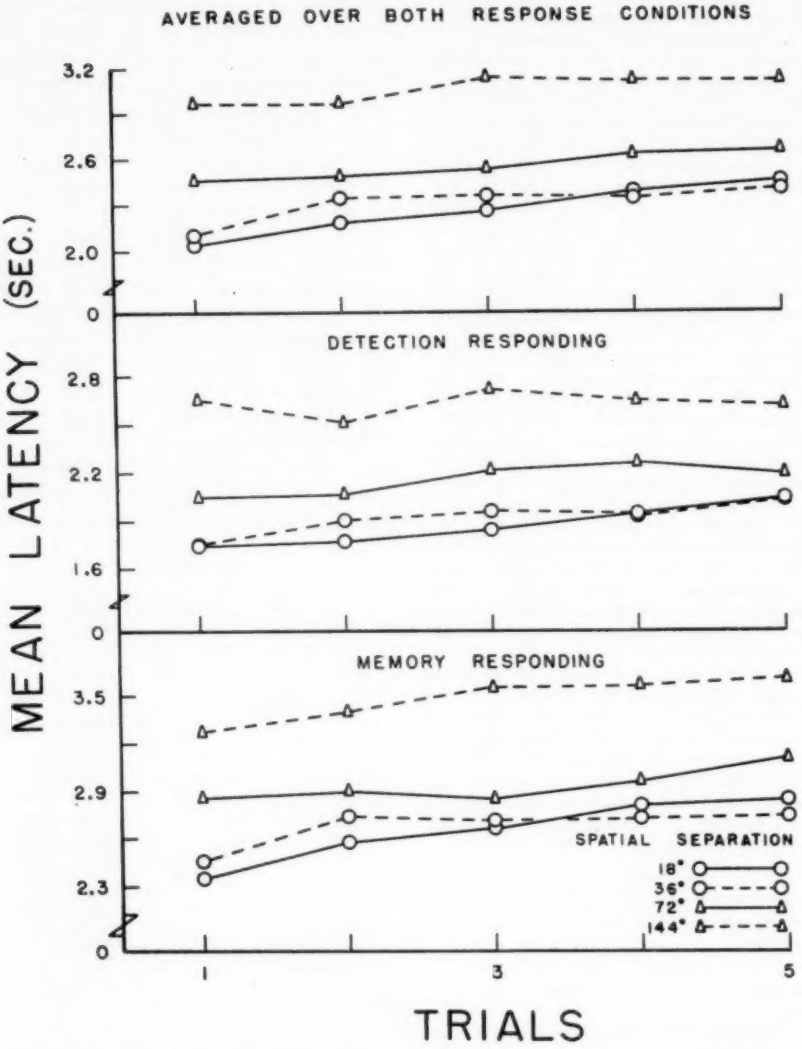

FIG. 1. Mean response latency as a function of the two types of response complexity and the four separations of stimulus sources: Exp. I.

complexity on decrement are sound, significant interactions should be expected. Less decrement over trials should be found for large spatial separations and for memory responding. The results of the analysis of variance in Table 1 show that the hypothesized interaction effects were not confirmed. All main effects, however, were significant.

Experiment II

The activationist hypothesis does not clarify the dimensional characteristics of stimulation that influence performance (Adams et al., 1961), and it could be that our operations in Exp. I did not define those particular characteristics of stimulation that deter vigilance decrement. The negative results of Exp. I may have resulted from the manipulation of *amount* rather than the *variety* of stimulation that has been found to be important by other investigators. Sharpless and Jasper (1956) found that stimulus variety induced alertness in the cat, and McGrath (1960) found that visual detection was heightened by varied auditory stimulation such as music, but not auditory white noise. Experiment II was performed to see if *change* in the pattern of head and eye movements could influence vigilance decrement. The plan of the experiment was to establish a pattern of visual observing by procedures analogous to operant reinforcement techniques and then, at a point in a session, require a change in the pattern.

Method

Apparatus.—The same task was used, but it was modified by mounting a small but distinctive neon cue light on top of each stimulus source. In addition to programing critical signals, whose duration was shortened to 2 sec. in this experiment, the tape input mechanism also programed the cue lights from source to source at 2-sec. intervals. The *S* was instructed always to watch the cue lights as they changed from source to source because the signal, whenever it occurred, would be coincident with an illuminated cue light. One way of viewing the modified task is in operant reinforcement terms where visually following the changing cue lights is a response class that is rewarded when *S* sees a number as it first comes on and has the success of an early detection. By faithfully following the cue lights, *S* tends to optimize his rewards and his performance. In this fashion, we sought to shape a pattern of observing responses that could be subjected to change.

Experimental procedures.—The 144° spatial separation was used for all groups to insure an ample observing response. Detection responding was used for all groups. One type of observing response was repetitive (R) where the cue light moved regularly, e.g., Sources 1, 2, 3, 4, 3, 2, 1 . . . etc., so that *S* scanned regularly back and forth. The other type of observing response was unsystematic (U) where the occurrence of cue lights was random with the restriction that the relative frequency at a source be the same as the R condition. Except for duration being reduced to 2 sec., the occurrences of critical signals on the input tapes were the same as in Exp. I. The shorter signal duration was used to place greater premium on close following of the cue lights because *S* would soon become aware that he would stand a chance of missing signals if he did not follow the cue lights closely. The length of the session was again 2.5 hr., so the shortening of signal duration slightly increased the length of intersignal intervals.

Four independent groups of 15 *S*s each were used. All *S*s first had a practice session, and on a subsequent day had a criterion session. Group RU had R scanning throughout except on Trial 3 of the criterion session when it was changed to U scanning. Group R was a control group for Group RU and had R scanning throughout both sessions. If the change of scanning pattern for Group RU influenced vigilance decrement, it would be detectable by comparison with Group R which had no change. As a check on the type of scanning and the effect of its change, another group, UR, had U scanning except when it was changed to R on Trial 3 of the criterion session. Group U was the control for UR, and had U scanning for both sessions.

Subjects.—The 60 *S*s were university male undergraduates who were paid for their participation. Assignment to groups was random.

Results

As in Exp. I, percent detection was high (98%), and the basic score for *S* again was the mean of his response latencies on a trial. A plot of group means as a function of trials is shown in Fig. 2. Group R, with repetitive visual scanning, had lower mean latencies than Group U with unsystematic scanning, but both groups had decrement. These apparent trends were confirmed in a mixed analysis of variance (Lindquist, 1953, p. 267) where type of scanning (U and R) was the between-*S*s variable, and trials the within-*S*s variable. Type of scanning gave an F ratio ($df = 1/28$) of 19.65 (error $MS = .17$) which was significant at less than the .01 level. Trials had an F ($df = 4/112$) of 9.11 (error $MS = .009$) which also was significant at less than .01. The interaction between these two main effects gave an F ($df = 4/112$) of .44 (error $MS = .009$) which lacked significance at the .05 level.

Trial 3 was the point of variation in scanning for Groups UR and RU, and Fig. 2 shows that the effect of the change was to shift a Trial 3 mean in the direction of its control group.

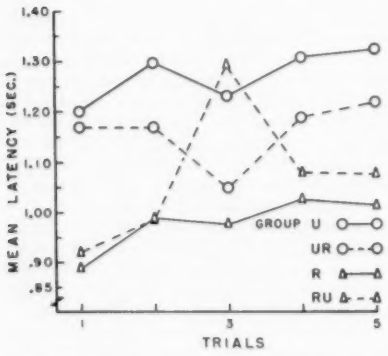

FIG. 2. Mean response latency for the four groups of Exp. II.

The t test for independent measures was used on Trial 2 latencies between Groups RU and R, and Group UR and U, to establish comparability of groups before the point of change. The t ratio for the mean difference between RU and R was .05 (error $MS = .004$), and for the difference between UR and U was 1.64 (error $MS = .006$). Both t's lacked significance. Groups RU and U, and Groups UR and R, were then evaluated with the t test at Trial 3, the point of change. The t for the difference between RU and U was .86 (error $MS = .005$), and UR and R was 1.05 (error $MS = .005$). Neither of these two t's was significant at the .05 level. To check if there was a persisting effect in the trial that followed, comparisons of Groups RU and R, and Groups UR and U, were made on Trial 4. The t for the difference between RU and R was .74 (error $MS = .005$), and the t for the comparison between UR and U was 1.50 (error $MS = .006$). Neither t was significant at the .05 level. For all of these t tests, $df = 1/28$.

Interpretation of these statistics must be founded on evidence that the cue lights changed observing behavior. First, cue lights produced lower mean latencies than when they were not used, presumably because *S* was directed where to look and was often reinforced when he looked there. Detection responding with 144° separation in Exp. I is essentially the same task as given Groups U and R, but without cue lights. Group U had almost twice the speed of responding as the corresponding condition in Exp. I, and Group R was almost three times faster. A second and more direct basis for inferring that cue lights influenced the observing response is shown in Fig. 3, where mean response latency to subsets of stimulus

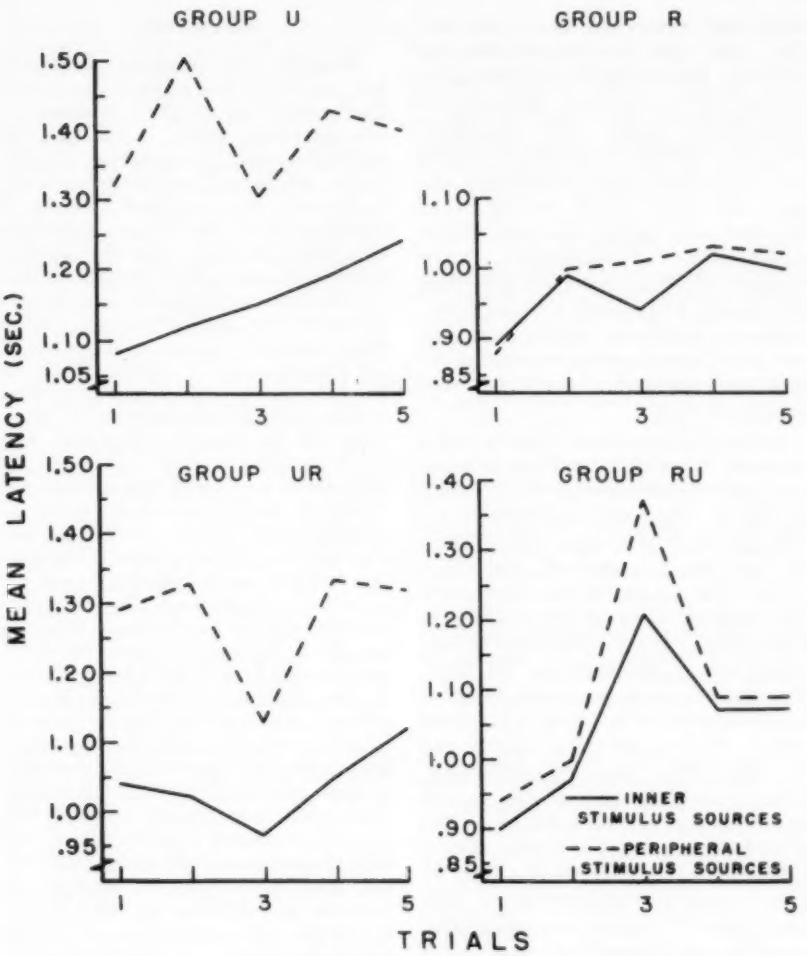

FIG. 3. For the four groups of Exp. II, mean response latency for the inner two and the outer two stimulus sources.

sources is presented. For each *S* a latency score was computed as the mean response time on a trial to signals of the two inner stimulus sources, and similarly another latency score was computed for signals on a trial for the two peripheral stimulus sources. Figure 3 shows the trial means for each group. If the observing response was under close control and *S* was systematically following the cue lights, then equal attentiveness to all sources is expected. Figure 3 shows that the expectation was realized for Group R. Not only was their performance level

high, but it was the same for both the inner and peripheral stimulus sources. Less success was achieved by Group U which had performance noticeably poorer for peripheral sources. The larger excursions of head and eye movements apparently did not fall under close control when eliciting cue lights changed unsystematically. Nevertheless, some measure of overall success was evident for Group U because its mean performance level was better than that for the corresponding condition of detection responding and 144° separation in Exp. I.

Group R and Group U each had a separate three-way analysis of variance performed on their data shown in Fig. 3. There was one measure per cell, and trials was one variable, inner vs. peripheral a second, and *S*s a third. For Group R the trials variable had an F ratio ($df = 4/56$) of 6.50 (error $MS = .014$) which was significant at less than the .01 level. The inner vs. peripheral variable gave an F ratio ($df = 1/14$) of 2.44 (error $MS = .009$) which failed significance at .05. The interaction between these two main effects also lacked significance at the .05 level ($F = .70$; $df = 4/56$; error $MS = .010$). Group U had trials significant at less than .01 ($F = 3.76$; $df = 4/56$; error $MS = .025$) and also the inner vs. peripheral variable significant at less than .01 ($F = 38.58$; $df = 1/14$; error $MS = .053$). The interaction between main effects was significant between the .05 and .10 levels ($F = 2.16$; $df = 4/56$; error $MS = .032$).

The plots for Groups UR and RU also are shown in Fig. 3. Change in the patterning of cue lights influenced performance at both inner and peripheral sources, and in directions consistent with those for Groups U and R.

Discussion

Molar behaviorism concerns itself with S-R operations, and physiology occupies itself with either the effect of stimuli on internal bodily states or the effect of bodily states on overt behavior. Physiological psychology, on the other hand, is directed toward completing the arc of lawfulness from stimulus, to under-the-skin states, to molar behavior. A more substantial network of empirical laws can be achieved by coordinating the laws of psychology and physiology, and a firmer scientific footing is gained. In the two experiments reported here, we sought to move in this direction by trying to link characteristics of the task, properties of the reticular formation, and overt vigilance behavior. Our results were negative. Neither the amount nor variety of stimulation induced by manipulating head and eye movements, nor response complexity, influenced vigilance decrement as we had predicted from the activationist hypothesis. An impressive number of facts for vigilance behavior casually fit the general framework of the activationist hypothesis, but the absence of operational definitions for the type of stimuli, as well as the characteristics of each stimulus class, gives the hypothesis little predictive capability for measures of molar behavior (Frankmann & Adams, 1962). Our two experiments suggest that the specification of relevant stimuli and their properties is hardly straightforward, and they emphasize the problems of operationally defining variables of the hypothesis well enough for it to be thoroughly tested and graduated into a substantive psychophysiological law.

There are several plausible reasons why our tests of the activationist hypothesis failed, and they all illustrate the difficulties of operational definition that must be resolved before the hypothesis can be accepted or rejected. Perhaps our manipulations did not induce differential stimulation of the reticular formation. Adams, Stenson, and Humes (1961) found that number of stimulus sources did not exert a discernible effect on vigilance decrement, and they conjec-

tured that increasing the number of stimulus sources might not have produced an increment in stimulation at the reticular formation. The same reasoning might be applied here. Or, perhaps the operations we used did not sufficiently arouse the reticular formation. We did not have direct physiological measurement of the impact at the reticular formation of these stimuli generated by head and eye movements, but vigorous reticular activation from these operations is certainly an expectation from our current physiological knowledge. Also, the operations of immediate memory might be questioned as a way of inducing cortical-centered stimuli, but the task of keeping the most recent four numbers in mind has a reasonableness about it for a central source of activation. Even if our selection of stimulus classes is granted, we still may not have uncovered manipulations for the relevant dimension of stimuli. Both the memory and the visual scanning tasks in Exp. I may have become repetitive rather early, and decrement could have occurred as a function of adaptation to the steady state of internal stimuli, as Scott (1957) has suggested. Early adaptation may have been a factor in Exp. II where we explicitly tried to introduce variety in our key stimuli. Sustained stimulus variety, whatever its operational definition might be, may prove to be the key to human alertness. Possible directions for inferring relevant dimensions of stimulus variety for visual alertness is the study by McGrath (1960) where meaningful audio stimulation such as music improved detection, and the studies by Adams and Boulter (1960) and Adams, Stenson, and Humes (1961) where four-choice decision responding improved monitoring behavior. McFarland, Holway, and Hurvich (1942) found that a brief interlude of stretching and conversation with *E* eliminated decrement in a measure of visual threshold.

The typical vigilance experiment, where marked decrements are found in prolonged monitoring behavior, differs from those presented here in that near threshold stimuli occur at a single stimulus source and percentage of signals detected is the measure. However, there is nothing that would seem to preclude our task with its latency measure as a sound vehicle for testing the hypothesis. Decrement was uniformly obtained but we were unable to control it by manipulating sources of stimulation.

Summary

The activationist hypothesis contends that environmental and internal sources of stimulation, working through the reticular formation of the brain, are sources of human alertness. Two experiments were performed to identify stimulus determinants of vigilance decrement in a complex visual monitoring task. The *S*'s task was to detect a two-digit number as rapidly as possible when it appeared at any one of four stimulus sources. Response latency was the measure of performance.

Experiment I sought to manipulate response-produced stimulation arising from the stimulation induced by head and eye movements and immediate memory. Amount of head and eye movements was defined in terms of the spatial separation of sources, and the physiological basis of its stimulation value for alertness was hypothesized to be proprioception and visual collaterals in the reticular formation. Immediate memory was defined as a requirement to remember the four numbers that last appeared at the sources and, when a number appeared, to respond with respect to the last number that had appeared at that source. Immediate memory was hypothesized to provide inputs to the reticular formation from cortical areas. No effects of these variables on vigilance decrement were found. Experiment II asked if the negative results of Exp. I were related to amount rather than variety of stimulation being manipulated. Experiment II introduced variety in head and eye movements by training the pattern of visual observing responses and then changing the pattern on a trial. No effects on vigilance decrement were found. Problems of operationally defining the activationist hypothesis were discussed.

References

Adams, J. A., & Boulter, L. R. Monitoring of complex visual displays: I. Effects of response complexity and intersignal interval on vigilant behavior when visual load is moderate. *USAF CCDD tech. Note*, 1960, No. 60-63.

ADAMS, J. A., STENSON, H. H., & HUMES, J. M. Monitoring of complex visual displays: II. Effects of visual load and response complexity on human vigilance. *Hum. Factors*, 1961, **3**, 213–221.

BROADBENT, D. E. *Perception and communication*. New York: Pergamon, 1958.

COOPER, S., DANIEL, P. M., & WHITTERIDGE, D. Muscle spindles and other sensory endings in the extrinsic eye muscles: The physiology and anatomy of these receptors and of their connections with the brainstem. *Brain*, 1955, **78**, 564–583.

FISKE, W. Effects of monotonous and restricted stimulation. In D. W. Fiske and S. R. Maddi (Eds.), *Functions of varied experience*. Homewood, Ill.: Dorsey, 1961. Pp. 106–144.

FRANKMANN, J. P., & ADAMS, J. A. Theories of vigilance. *Psychol. Bull.*, 1962, **59**, 257–272.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

LINDSLEY, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 44–105.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

MCFARLAND, R. A., HOLWAY, A. N., & HURVICH, L. M. Studies of visual fatigue. *Publ. Grad. Sch. Bus. Admin. Harvard U.*, 1942.

MCGRATH, J. J. The effect of irrelevant environmental stimulation on vigilance performance. Project on Human Factors Problems in Anti-Submarine Warfare, Technical Report No. 6, 1960, Office of Naval Research, Psychological Sciences Division.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

ROSSI, G. F., & ZANCHETTI, F. The brain stem reticular formation. *Arch. Ital. Biol.*, 1957, **95**, 199–435.

SAMUELS, I. Reticular mechanisms and behavior. *Psychol. Bull.*, 1959, **56**, 1–25.

SCOTT, T. H. Literature review of the intellectual effects of perceptual isolation. Report No. HR66, 1957, Department of National Defence, Defence Research Board, Canada.

SHARPLESS, S., & JASPER, H. Habituation of the arousal reaction. *Brain*, 1956, **79**, 655–680.

WHITTERIDGE, D. Central control of eye movements. In J. Field (Ed.), *Handbook of physiology*. Washington: American Physiological Society, 1960. Pp. 1089–1109.

(Received October 28, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 505-512

THE RELATIVE EFFICIENCY OF SEVERAL TRAINING METHODS AS A FUNCTION OF TRANSFER TASK COMPLEXITY [1]

GEORGE E. BRIGGS AND JAMES C. NAYLOR

Ohio State University

In a recent review Naylor (1962) has suggested that the relative efficiency of part- and whole-task training schedules is a function of two variables: *complexity* and *organization*. In his taxonomy, Naylor defines complexity as a function of the *information-processing* demands imposed on *S* by each component of the task separately, while task organization is specified by the demands placed on *S* by the interactions or interrelations *between* task components. Thus, task complexity is an intracomponent characteristic, task organization an intercomponent characteristic.

Naylor (1962) has further suggested that the efficiency of a whole-task training method (as compared to a pure-part task method) will increase as task complexity is increased, i.e., there should be an interaction between training method and task complexity in terms of transfer performance. Therefore, those tasks which place greater intradimensional demands upon *S* should be learned best by whole practice, whereas those tasks having lesser intradimensional task requirements should show a decrease in the relative efficiency of the whole training method.

[1] This research was carried out in the Laboratory of Aviation Psychology and was supported by the United States Navy under Contract No. N61339-950, sponsored by the United States Naval Training Device Center, Port Washington, New York. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part for any purpose of the United States Government.

In tracking tasks, complexity is defined by the order of the machine dynamics through which *S* attempts to control. Thus, in a multidimensional tracking task, the dynamics of *each* dimension are what collectively determine total task complexity. More specifically, Birmingham and Taylor (1954) pointed out that an increase in dynamics, as from a simple positional (such as setting one's watch) to rate control (such as steering an auto), results in an increase in the information-processing demands placed on *S*, i.e., he must increase the use of derivative information in the tracking error signal to maintain stability in rate task control. In fact, Russell (1951) clearly showed that compared to his transfer function (an analytic expression of his information-processing functions) with a positional control, *S* will increase his damping (weighting of the first derivative of error in determining a control movement) by a factor of 25 when controlling through rate dynamics. Thus, a more simple transfer function is adequate for position control tasks, but rate dynamics require rate estimates by *S*, thereby increasing the complexity of his task.

Similarly, a further increase in dynamics from rate control to acceleration control results in an even more complex task dimension, since *S* now must provide a transfer function with the equivalent of two differentiations, one for each of the integral lags present in the control dimension.

The relationship between information-processing demands and control dynamics is discussed in detail by Birmingham and Taylor (1954).

The present study represents an investigation of the task complexity hypothesis. In addition to whole and pure-part task training methods, the present study included groups trained via either a progressive-part or a simplified-whole method.

METHOD

Apparatus.—A three-dimensional compensatory tracking system defined the skill task. The *S* observed three center-reading meters labeled "heading," "altitude," and "yaw," and he was instructed to maintain null readings on these meters as much of the time as possible. A single input signal was applied to all three dimensions in the form of a simple sinusoid of .03 cps. In the more complex versions of the task (see below) meters showing "altitude rate" and/or "heading rate" were provided. All meters were mounted on a single 10.5 × 19 in. panel which was located approximately 24 in. from *S*, who manipulated a single, three-dimensional control. The control device was a two-dimensional stick where the three control stick movements were left-right (heading), front-back (altitude), and clockwise-counterclockwise rotation of the head (yaw). All three control-display linkages conformed to population stereotypes.

An EASE analog computer provided the task dynamics, as illustrated in Fig. 1. As indicated, there were three levels of control complexity used in this study. Tracking proficiency was measured during each trial in terms of integrated absolute tracking error (average error) for each of the three dimensions separately.

Design.—Two independent variables were manipulated in a transfer of training paradigm: transfer task complexity (two levels) and method of training (four methods). The two levels of transfer task complexity were Level II and Level III of Fig. 1. The four training methods were whole-task training, pure-part training, progressive-part training, and simplified-whole task training. There were two groups of *S*s for each of the training methods (one for each of the transfer task complexity levels). The two whole-task groups tracked through Level II and Level III dynamics, respectively, for a total of 10 daily

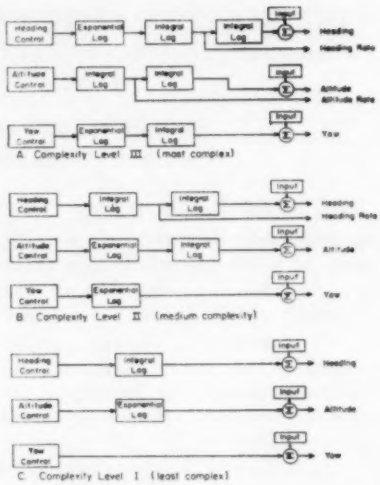

FIG. 1. System dynamics used to obtain various levels of task complexity.

sessions. The remaining six groups received 8 sessions on a training task and 2 sessions on the final transfer task. For three of the six groups the final transfer task involved Level II dynamics, while the other three groups experienced Level III dynamics upon final transfer.

Pure-part training involved four 35-sec. trials per session on each of three separate dimensions. Progressive-part training involved an initial three sessions on a pure-part schedule followed by five sessions during each of which *S* practiced four 35-sec. trials on each of the three possible pairings of the

TABLE 1

EXPERIMENTAL DESIGN OF TRAINING AND TRANSFER CONDITIONS

Group	Training Condition	Training Task	Transfer Task
1	Whole	Level III	Level III
2	Pure-Part	Level III	Level III
3	Progressive-Part	Level III	Level III
4	Simplified	Level II	Level III
5	Whole	Level II	Level II
6	Pure-Part	Level II	Level II
7	Progressive-Part	Level II	Level II
8	Simplified	Level I	Level II

Note.—See Fig. 1 for a definition of Complexity Levels I, II, and III.

heading, altitude, and yaw dimensions. Of the two groups which experienced the simplified-whole training method, one trained for eight sessions on Level I dynamics (see Fig. 1) and then transferred to Level II, while the other group trained on Level II and transferred to Level III. Table 1 summarizes the training and transfer conditions.

Subjects and procedure.—A total of 144 male undergraduates participated in this study. Assignment to the eight groups was based on randomization procedures with the restriction that groups be filled approximately equally throughout the data collection period. At the end of this period there were 18 *Ss* per group. The *Ss* were reimbursed $10.00 for their time.

There were 12 35-sec. tracking trials during each of the 10 sessions. Practice occurred in four-trial blocks with 25 sec. rest between trials within a block and 1 min. rest between blocks. Scoring occurred over the final 30 sec. of each trial to avoid the initial transients in tracking performance; thus, a block of four trials served as the unit of data analysis and represents a behavioral sample of 2 min. for each *S*. During the 12 trials of the initial session, all *Ss* controlled the task *without* the .03-cps input signal; that input was present during all remaining trials. The purpose of this procedure was to facilitate the initial acquisition of skill. All three levels of task dynamics provide inherently unstable systems, and thus the system tends to go out of control quite rapidly for the skill novice; removing the input markedly improved *S*'s chances of "keeping control" over the system during the initial training trials as it provided a less difficult task.

All *Ss* were instructed to keep the meter displays representing heading, altitude, and yaw at their null position at all times, if possible. It was emphasized that control over all three task dimensions was required, and *S* was reminded of this if his performance indicated that he was paying particular attention to one or two displays to the detriment of the other dimension(s). The *S* was informed also of the usefulness of the rate display (if it were present, see Fig. 1).

The performance metric, average error, is actually the average deviation of the error amplitude distribution generated by *S* during each tracking trial. Therefore, the metric describes *S*'s intratrial variability in tracking accuracy. The score was recorded originally in voltage units; however, following data collection all scores were transformed to units of inches of arc, thereby providing a scale matching that of *S*'s tracking display.

Results

The results from both the training and the transfer sessions are summarized in Fig. 2 (Groups 1–4) and Fig. 3 (Groups 5–8). The tracking error scores were combined for all three dimensions and averaged for each group. In order to provide comparable data points, it was necessary to derive daily averages for the training sessions, i.e., the pure-part and the progressive-part groups experienced all three tracking dimensions during each of the eight training sessions; however, during transfer it was possible to derive comparable data points for four-trial blocks. The use of combined error scores is justified here, since the performance functions on the three dimensions are quite similar in relative rank order and no apparent artifacts are introduced by dealing with "total task" performance in either the training or the transfer sessions.

Training.—Several points are of general interest from the training data of Fig. 2 and 3. First, as expected, tracking performance is best during Sessions 1 through 8 for that whole-task group, Group 8, which experienced the least complex system (Level I), and the least proficient performance was attained by Group 1 which experienced the most complex

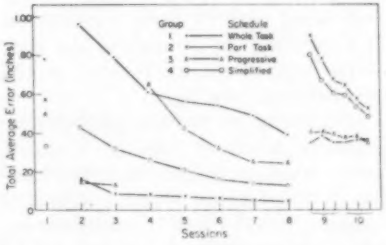

FIG. 2. Training and transfer performance levels for Groups 1–4. (Transfer to Task Complexity Level III.)

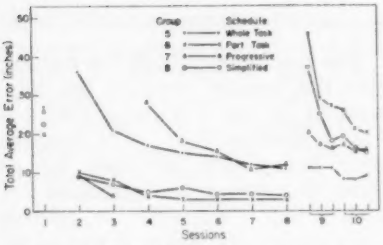

FIG. 3. Training and transfer performance levels for Groups 5–8. (Transfer to Task Complexity Level II.)

whole task (Level III). Both Groups 4 and 5 trained on the same whole task (Level II), and their performance levels (*a*) are not significantly different one from another ($P > .05$) and (*b*) fall intermediate to those of Groups 1 and 8.

It may be recalled that no input signal was employed during Session 1 for any group; thus, these data are set apart in Fig. 2 and 3. Also, it may be recalled that Groups 3 and 7 transferred from a pure-part schedule to practice on pairs of task dimensions on Session 4; therefore, the figures show a discontinuity between Sessions 3 and 4 for Groups 3 and 7. It is interesting to note on Session 4 and thereafter that Group 3 surpasses the whole-task group, Group 1, whereas Group 7 never does exceed the performance of Group 5, even though Group 7 experienced only two tracking dimensions while Group 5 was confronted with all three dimensions. This may well be a result of the very simple yaw dynamics present in the Level II task (see Fig. 1).

In addition to the comparable performance levels of Groups 5 and 7 late in training, it may be noted from Fig. 3 that Groups 6 and 8 achieved similar "total task" performance. Again, this is probably a result of the very simple task dynamics encountered by Group 8 during training. The more complex dynamics experienced by Groups 1–4 would account, therefore, for the clear separation of performance levels for these groups during the latter training sessions (see Fig. 2).

An analysis of variance was applied to the data of Groups 1–3 and Groups 5–7 on Session 8. Groups 4 and 8 were not included, since their task dynamics differed from those of the companion groups. The results of the analysis indicated that training methods ($F = 78.6$; $df = 2/102$), task complexity ($F = 85.2$; $df = 1/102$), and the Methods × Complexity interaction ($F = 28.0$; $df = 2/102$) were statistically significant at $P < .05$. The interaction, illustrated in Fig. 2 and 3, shows that the greatest performance difference between the Level II and the Level III tasks occurred with the whole-task groups, while the progressive-part groups showed less influence of the task complexity variable, and no difference ($P > .05$) was found between Groups 3 and 7. The only other group difference which did not attain statistical significance at $P < .05$ on Session 8 was that between the whole and progressive-part groups on the Level II task dynamics (Groups 5 and 7 as mentioned earlier).

Finally, it may be noted from Fig. 2 and 3 that all groups had approached asymptotic performance levels at or before transfer. This was especially the case for the two part-task practice groups (Groups 2 and 6). If anything, Groups 2 and 6 were overtrained in eight sessions.

Transfer.—It would appear from Fig. 3 that during the initial block of transfer trials Groups 6 and 8 performed at a level inferior to Group 5 on Session 2. It should be recalled that the data point for Group 5 on Session 2 is the average for three blocks of trials, and when performance for Group 5 on the first *block* of Session 2 is compared to that of

Groups 6 and 8 on the first block of Session 9, the result is not quite as unexpected. The transfer index $(C_i - E_t/C_i - C_t) \times 100$ was used for this comparison, where C_i is the performance of Group 5 on the first block of Session 2, C_t is the performance of the same group on the first block of Session 9, and E_t is the performance of Group 6, 7, or 8 on the latter block of trials. This index of transfer is particularly appropriate when the raw data are error scores, as in the present case, and the index describes relative improvement in performance as a function of training.

The same index was applied to the data of Groups 1–4 and the results for all groups were: part training = 2% transfer to Complexity Level II, 31% to Complexity Level III; progressive-part training = 75% transfer to Level II, 94% to Level III; and simplified training = 28% transfer to Level II, 44% to Level III. One may note that (*a*) Group 3 attained a very high level of performance during transfer (94%), while Groups 2, 4, 6, and 8 attained rather poor relative transfer performance levels, and (*b*) transfer performance was relatively higher on the more complex task than on the less complex task. It is important to note that these transfer data are expressed in relative terms, as indicated above. If examined in terms of *absolute* transfer performance levels, the pattern of results as listed is partially reversed.

This latter point may be illustrated by reference to an analysis of variance which was performed on the data of all eight groups over the entire six blocks of transfer trials. The results are summarized in Table 2 where it is apparent that, as was the case with the training data, there are significant differences among the training methods and between the two levels of task complexity. Further, and of direct relevance to the above statement on absolute transfer performance levels, the interaction of training methods by task complexity is statistically significant at $P < .05$. Figure 4 contains the group averages which define this interaction. There it is obvious that considerably greater absolute differences occurred between the whole-task group (Group 1) on the more complex transfer task and the groups trained by either pure-part or simplified-whole methods (Groups 2 and 4) than between the comparable groups on the less complex transfer task (Group 5 vs. Groups 6 and 8). However, again it should be emphasized that *relative* comparisons between control (whole training) groups and the experimental groups show better transfer performance on the more complex task than on the less complex task for the latter groups.

The other statistically significant interactions indicated in Table 2 are of interest. The Transfer Blocks × Training Methods interaction indicates a difference in the slopes of the transfer performance curves for the several groups. By reference to Fig. 2 and 3 it may be seen that the slopes for the whole task and progressive-part task groups are very slight, whereas those for the part and

TABLE 2

ANALYSIS OF VARIANCE AS APPLIED TO THE TRANSFER TRIAL BLOCKS OF SESSIONS 9 AND 10 FOR ALL GROUPS

Source	*df*	*MS*	*F*
Training methods (M)	3	.764	16.6*
Task complexity (C)	1	5.410	117.6*
M × C	3	.132	2.9*
Ss within methods (*Ss*/M)	136	.046	
Transfer blocks (B)	5	.122	40.7*
B × M	15	.030	10.0*
B × C	5	.007	2.3*
B × C × M	15	.003	1.0
B × *Ss*/M	680	.003	

* $P < .05$.

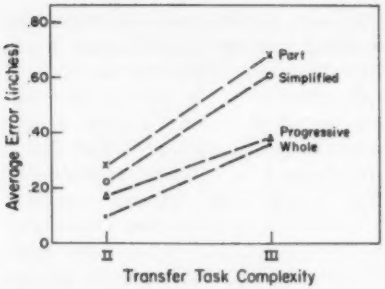

FIG. 4. Transfer task performance averaged over Sessions 9 and 10 for all groups.

simplified-whole groups are rather steep. Thus, while initial transfer performance for the latter two groups was relatively poor, their progress following initial transfer was quite rapid.

The Transfer Blocks × Task Complexity interaction is more difficult to see in Fig. 2 and 3. However, it is a result primarily of the especially high rates of improvement by Groups 2 and 4 on the more complex task. It follows, therefore, that not only was relative transfer higher for these groups on the more complex task, but also their rate of approach to the performance level of the control group (Group 1) was greater than that for the comparable groups on the less complex system.

Following the analysis of variance, summarized in Table 2, a Duncan multiple range test (Duncan, 1955) was applied to the data for all groups on the first block of transfer trials and on the last block. A summary of these comparisons is provided in Table 3. The upper half of each section of Table 3 represents comparisons made with the initial block of transfer trials, while the lower half is from the last block of transfer trials. These results show that whereas there were differences among groups early in transfer, the relatively short transfer experience with the whole task eliminated all group differences; therefore, the effects of training method were quite transitory.

TABLE 3

GROUP COMPARISONS OF TRANSFER PERFORMANCE ON BLOCK 1 (UPPER PART) AND ON BLOCK 6 (LOWER PART) FOR ALL GROUPS

Cond.	Level III				Level II			
	W	P	P-P	S	W	P	P-P	S
W		.05	*ns*	.05		.05	*ns*	.05
P	*ns*		.05	*ns*	*ns*		.05	*ns*
P-P	*ns*	*ns*		.05	*ns*	*ns*		.05
S	*ns*	*ns*	*ns*		*ns*	*ns*	*ns*	

Note.—Cell entries indicate statistical significance at $P < .05$ and at $P > .05$ (*ns*). W, P, P-P, and S identify whole, part, progressive-part, and simplified-whole methods.

DISCUSSION

Several findings of the study are of particular interest. First, the relative transfer indices are higher for all three experimental groups on the more complex task. Using this measure of method efficiency, one finds an indication that the part-training methods are actually more efficient when the task is highly complex —a denial of the Naylor hypothesis. Second, using absolute transfer performance (see Fig. 4), one finds that only the progressive-part schedule showed an improvement relative to the whole method. These data lead to the conclusion that both the pure-part and the simplified-whole methods become less efficient training techniques for highly complex tasks and that the whole and some schedule of progressive-part methods emerge as more satisfactory training techniques as task complexity is increased. This substantiates the hypothesis that task complexity and training method interact, with whole training becoming relatively more efficient as complexity is increased. Since the validity of the hypothesis therefore de-

pends upon the measure of transfer that is used, it is not possible to accept or reject without ambiguity.

One possible source for this ambiguity lies in the second task variable mentioned by Naylor (1962). It is likely that an increase in component complexity results in a corresponding increment in the organizational demands of the task. With the three-dimensional task used here, the primary organizational requirements imposed on *S* consisted of time-sharing demands, i.e., he had to develop an efficient pattern of attention alternation among the three displays and avoid undue emphasis on a single system dimension (component). Thus, as task complexity is increased, *S* must coordinate his responses to the several task dimensions more exactly to avoid large errors in one or another of the dimensions.

A comparison illustrating the importance of time sharing is that between the simplified-whole, the progressive-part, and the pure-part training methods. If increases in complexity of the system components influence task organization, then the pure-part method should show less effect of task complexity, while the progressive-part and the simplified-whole methods should be affected to a greater extent. This prediction was substantiated by the data: there is no statistically significant difference between Groups 2 and 6 (pure-part) during the training sessions; however, both Groups 3 and 7 (progressive-part) and Groups 4 and 8 (simplified-whole) differ markedly (see Fig. 2 and 3).

Perhaps one of the more intriguing findings related to time-sharing concerns the relationship between the whole and the progressive-part groups during transfer. It would appear that time-sharing can be learned quite well by the progressive-part method, as there were no significant differences between the whole and progressive-part groups at either level on Transfer Block 1 (Table 3). Since, at the most, only practice on pairs of task dimensions was given in the progressive-part method, the acquisition of time-sharing skill may not require practice on the total task.

It would appear, then, that the time-sharing demands of multidimensional control tasks represent a potent variable determining the relative efficiency of training methods. However, this variable accounts for only part of the transfer results. If time-sharing requirements were the only variable of importance, then those groups trained via a simplified-whole method should have been superior to the pure-part groups during transfer. It may be noted from Table 3 that these groups did not differ significantly at $P < .05$; thus, the apparent superiority of Groups 4 and 8 over Groups 2 and 6, respectively, in Fig. 2 and 3 is only apparent—not real. This was a completely unexpected finding of the present research. It was expected on the basis of an earlier analysis by Briggs and Waters (1958) that a simplified-whole training method would provide superior transfer performance compared to that following pure-part training.

Also, Briggs, Fitts, and Bahrick (1958) had previously demonstrated that greater amounts of training resulted in greater amounts of transfer for a simplified training procedure. Since the amount of training provided in this study was greater than that used in the Briggs et al. study, large amounts of transfer for the simplified groups were anticipated.

Therefore, it may be that the benefit accrued by the simplified-whole groups during training (the acquisition of time-sharing skills) was offset in comparison with the pure-part groups by their training on the specific dynamics of the transfer task. Similarity of training and transfer tasks emerges, therefore, as a second major determinant of transfer performance.

It is concluded that some form of progressive-part training will be superior to methods such as pure-part and simplified-whole for the acquisition of skill in a complex, multidimensional task, since the progressive-part method utilizes a training task of high similarity to the transfer task and it also provides an opportunity to develop efficient time-sharing behavior.

Finally, attention is directed to the fact that practice on the whole task during transfer rather quickly eliminated, at least statistically, the differential effects of training methods (see Table 3). Since whole-task training resulted in numerically superior performance throughout training, the only argument for the pure-part or simplified-whole methods rests upon the potential savings in training time on the whole task itself following training. This argument is tenable only to the extent that (*a*) a minimum of whole-task training can equate the effects of the methods in terms of final performance, and (*b*) during transfer following pure-part or simplified-whole training the initial proficiency of *S* is able to meet some criterion of operating safety, i.e., it is conceivable that even after extended training *S* could "lose control" of a complex system and thereby violate safety margins for himself and the system.

The first point above was met in the present study by the very rapid progress of the pure-part and simplified-whole groups during transfer; however, their original transfer performance was very poor, and the pure-part groups, in particular, probably would have been judged "dangerous" had their training and transfer tasks been in a real-life vehicular system.

Summary

The relative efficiency of four training methods (pure-part, progressive-part, simplified-whole, and whole task) in the acquisition of skill was investigated in a three-dimensional tracking task. Two levels of transfer task complexity were used. There were eight daily training sessions followed by two transfer sessions on the whole task.

The results showed that training via the whole and the progressive-part methods resulted in statistically equivalent performance during transfer for both levels of transfer task complexity. Both the whole and the progressive-part methods resulted in transfer performance which was statistically superior to that of *S*s trained via either the pure-part or the simplified-whole method. Further, as task complexity was increased, the whole and the progressive-part training methods increased the absolute (but not the relative) superiority of transfer performance compared to that with pure-part and simplified-whole methods.

Two factors emerged from a logical analysis of the results to explain the superiority of the whole and the progressive-part methods: first, the transfer (whole) tasks required rather efficient time-sharing skills and these could not be acquired under the pure-part method since, by definition, *S* experiences only one dimension of the task at a time under this training method, and second, training and transfer task similarity was not as high for the simplified-whole groups as it was for those *S*s trained under the progressive-part method; thus, even though the former *S*s could acquire the necessary time-sharing skills, they were "penalized" by training on a task which bore less similarity to the transfer task than was the case for the latter *S*s.

References

Birmingham, H. P., & Taylor, F. V. A human engineering approach to the design of man-operated continuous control systems. *USN Res. Lab. Rep.*, 1954, No. 4333.

Briggs, G. E., Fitts, P. M., & Bahrick, H. P. Transfer effects from a single to a double integral tracking system. *J. exp. Psychol.*, 1958, **55**, 135–142.

Briggs, G. E., & Waters, L. K. Training and transfer as a function of component interaction. *J. exp. Psychol.*, 1958, **56**, 492–500.

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

Naylor, J. C. Parameters affecting the relative efficiency of part and whole practice methods: A review of the literature. *USN Train. Dev. Cent. tech. Rep.*, 1962, No. 950-1.

Russell, L. Characteristics of the human as a linear servo-element. Unpublished master's thesis, Massachusetts Institute of Technology, 1951.

(Received October 30, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 513–518

RECRUITMENT, LATENCY, MAGNITUDE, AND AMPLITUDE OF THE GSR AS A FUNCTION OF INTERSTIMULUS INTERVAL [1]

WILLIAM F. PROKASY, JAMES T. FAWCETT [2] AND JOHN F. HALL

Pennsylvania State University

White and Schlosberg (1952) and Moeller (1954) report that, within the ranges of 0 and 5 sec., the optimum interstimulus interval (ISI) in GSR conditioning is approximately .5 sec. when magnitude of response is employed as the dependent variable. Bierbaum (1955), in contrast, reports a greater response magnitude near 3 sec.

While it is not clear why his ISI function should diverge from that obtained by White and Schlosberg (1952) and Moeller (1954), Bierbaum (1955) reports, in addition, that recruitment (time elapsing between the onset of a GSR and its maximum) increases with increases in the ISI. Unpublished data from our laboratory suggest, further, that both GSR latency and recruitment modify across extinction trials.

It is the purpose of the present study (*a*) to examine further GSR conditioning as a function of ISI and (*b*) to relate the independent variable (ISI) to three GSR attributes other than magnitude: latency, recruitment, and amplitude. The distinction between magnitude and amplitude first was made by Humphreys (1943) and more recently was discussed by Hilgard (1951, p. 528). These authors employed the word *magnitude* to refer to a mean based upon all trials, including those which resulted in no measurable response, while they adopted the word *amplitude* to refer to means derived only from those trials on which a response occurred.

[1] This study was supported by an NSF grant (G-7463) to the senior author.

The authors gratefully acknowledge the aid of Herbert Krauss and Barry Lively for their aid in transforming the data to log conductance units, and of the staff at the computer center of Pennsylvania State University for making the IBM 650 available to us.

[2] Now with the Peace Corps, Washington, D. C.

METHOD

Subjects.—The *S*s were 129 men and women enrolled in introductory psychology who volunteered with the knowledge that shock would be employed. Of these, 23 were lost due to equipment malfunction, adaptation to shock, or *E*'s error.

Apparatus.—The GSR was measured by means of the Fels dermohmeter. The recording electrodes were .75-in. zinc electrodes set at the base of Plexiglas cups of .125-in. depth. A zinc oxide electrode paste filled the cups and made contact with *S*. The cups were placed approximately 1 in. apart (center to center) on *S*'s left palm with a constant current of 70μa. transmitted between them. Resistance changes were recorded on an Esterline-Angus milliammeter operating at a paper speed of 24 in. per min.

The CS, a 76-db. (re .0002 dynes/cm^2) 1000-cps tone, was presented by means of a Grason-Stadler twin oscillator through Permoflux PDR-8 earphones, and was superimposed on a constant background of white noise generated through the earphones at 52 db. The UCS was a shock administered by an ac variac through .5-in. copper electrodes taped to *S*'s right index finger. CS and UCS duration were controlled through Hunter-Brown interval timers.

Procedure.—The *S*s were assigned, randomly, to one of five treatment conditions: 0-, .5-, 1-, 3-, or 5-sec. ISI. CS durations for the five conditions, respectively, were: .2, .5, 1, 3, and 5 sec. UCS duration was .2 sec. and began with the termination of the CS in all except Group 0, in which the CS and UCS overlapped.

The *S* was seated in a sound-shielded room (6 × 10 × 6.5 ft.) which was illuminated by a 75-w. incandescent bulb. After affixing GSR and shock electrodes, *E* indicated that shock would be administered in increasing amounts, beginning with a barely perceptible level, until it reached a level which *S* judged to be "highly annoying, but not painful." The mean voltage achieved was 58.7 (*SD* = 14.5). The *S* was then instructed that he would receive either the already experienced shock to the finger, a tone through the earphones, or both, but that there was no necessary relationship between the two events. The earphones were positioned and *E* returned to the outer chamber to initiate training trials. All *S*s received 20 CS-UCS pairings followed by 10 test (extinction) trials. The intertrial interval was varied unsystematically between 25 and 35 sec. Final *N*s in each group were: Group 0, 20; Group .5, 22; Group 1, 22; Group 3, 20; and Group 5, 22.

Response measures.—The measure employed on each test trial was the log of conductance change, as measured from the base at CS onset to the point of maximum pen deflection following any response initiation that occurred between 1 and 5.5 sec. after CS onset. To log conductance change was added a constant of 9, thus making all scores positive with a range of from .549 to 3. (Only 3 *S*s yielded scores below 1, and then on only several trials each.) In the event that the pen deflection failed to meet the criterion of a response, a score of zero was recorded. The minimum deflection recorded (as measured by a specially constructed template) was 25, 50, 100, 250, or 500 ohms, depending upon the dermohmeter sensitivity setting required to adjust appropriately to the extent of *S*'s UCR during shock adjustment.

Both magnitude and amplitude were recorded as the mean of log conductance change plus 9 on Test Trial 1 and on each of three blocks of three of the subsequent nine test trials. Magnitude was obtained by including the score of zero on trials in which no response occurred. The amplitude index did not include scores of zero. For example, if *S*'s scores on Test Trials 2, 3, and 4 were 2.5, 0, and 1.1, mean magnitude for those trials would be 1.2 while mean amplitude would be 1.8.

Latency was defined as the time elapsing between CS onset and the occurrence of the first response in the criterion range, while recruitment was defined as the amount of time elapsing between response initiation and response peak. The units of our template were in .03-in. increments which, in conjunction with the paper speed of 24 in. per min., permitted time measurement to the nearest 75 msec.

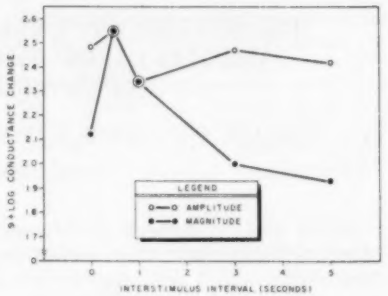

FIG. 1. Amplitude and magnitude as a function of ISI, measured on Test Trial 1.

RESULTS

First extinction trial.—Responses on the first extinction trial were employed as the most direct index of the influence of the ISI variable, as subsequent responses necessarily reflected the added operation of UCS omission. Mean magnitude and amplitude as a function of ISI are shown in Fig. 1. Amplitude means, based only upon those *S*s who responded, are obtained from *N*s of 17, 22, 22, 16, and 17 for Groups 0, .5, 1, 3, and 5, respectively.

Magnitude is greatest at .5 sec. and declines with longer ISI intervals, the decrease over ISI values of .5 to 5 sec. being significant at the .05 level

TABLE 1

LATENCY AND RECRUITMENT MEANS AND *SD*s ON TEST TRIAL 1 AS A FUNCTION OF ISI

ISI (Sec.)	*N*	Latency (Sec.)		Recruitment (Sec.)	
		Mean	*SD*	Mean	*SD*
0	17	2.48	.79	2.50	.72
.5	22	2.55	.48	3.14	1.04
1	22	2.58	.90	2.93	.90
3	16	3.11	1.22	3.23	.96
5	17	2.76	.98	2.87	1.82

TABLE 2

MEAN LATENCY, RECRUITMENT, AMPLITUDE, AND MAGNITUDE OVER THREE BLOCKS OF THREE TEST TRIALS AS A FUNCTION OF ISI

ISI (Sec.)	*N*	Amplitude			Magnitude			Latency (Sec.)			Recruitment (Sec.)		
		Trial Block			Trial Block			Trial Block			Trial Block		
		1	2	3	1	2	3	1	2	3	1	2	3
0	9	2.457	2.384	2.581	2.285	1.854	2.459	2.53	2.58	2.52	2.05	2.06	2.06
.5	19	2.506	2.375	2.343	2.459	2.197	2.136	2.48	2.54	2.49	2.15	2.09	2.12
1	18	2.451	2.400	2.391	2.415	2.223	2.203	2.47	2.58	2.75	2.61	2.39	2.36
3	17	2.409	2.305	2.404	2.201	2.085	1.926	3.17	3.31	3.32	3.23	2.84	3.02
5	18	2.402	2.367	2.290	2.374	2.230	2.096	2.89	2.95	3.41	2.93	3.00	2.73

Note.—Within-cell *SD*s varied from .18 to .50 for amplitude, from .24 to .80 for magnitude, from .27 to 1.14 for latency, and from .44 to 1.21 for recruitment.

($F = 3.98$, $df = 3/82$, error $MS = .593$). A similar analysis of amplitude did not result in significant differences ($F = 1.57$, $df = 3/73$, error $MS = .119$).

Table 1 summarizes latency and recruitment data on Test Trial 1. For neither recruitment ($F = 1.1$, $df = 4/90$, error $MS = 1.273$) nor latency ($F = 1.37$, $df = 4/90$, error $MS = .775$) were the differences among groups significant, although it can be noted that the longest latencies occurred in Groups 3 and 5.

Final nine extinction trials.—Data of all *S*s who responded at least once in each of three blocks of three extinction trials were employed in the analyses of variance of the four response measures. Table 2 summarizes the means for all groups while Table 3 provides a summary of the analyses of variance. Significant between-groups effects were obtained only with the recruitment and latency measures, in both instances a larger time value being associated with longer ISIs. The significant between-*S*s effect obtained with all measures attests to their reliability.

Both magnitude and latency changed across test trial blocks, with magnitude decreasing and latency increasing. There was a tendency for amplitude to decrease, suggesting that there was some decrease in the size of a response provided that one occurred.

TABLE 3

ANALYSES OF VARIANCE OF MAGNITUDE, AMPLITUDE, LATENCY, AND RECRUITMENT AS A FUNCTION OF ISI OVER THREE BLOCKS OF THREE TEST TRIALS EACH

Source	*df*	Amplitude		Magnitude		Latency		Recruitment	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between *S*s	80								
ISI	4	.080	.37	.572	.76	6.061	9.06***	8.879	6.91***
Error (b)	76	.216	3.86***	.752	4.82***	.668	2.58***	1.285	3.65***
Within *S*s									
Trials	2	.143	2.56	1.716	11.00***	.935	3.61*	.521	1.48
ISI × Trials	8	.056	1.00	.056	.36	.264	1.02	.204	.58
Error (w)	152	.056		.156		.259		.352	

* $P < .05$.
*** $P < .001$.

TABLE 4

MAGNITUDE MEANS AND *SD*s FOR ALL *S*s AS A FUNCTION OF ISI OVER THREE BLOCKS OF THREE TEST TRIALS EACH

ISI (Sec.)	*N*	Test Trial Block					
		1		2		3	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
0	20	1.411	.95	1.030	.98	1.116	1.11
.5	22	2.260	.58	1.960	.81	1.845	.95
1	22	2.183	.72	1.931	.92	1.915	.81
3	20	1.926	.89	1.892	.92	1.637	.98
5	22	2.007	.93	1.925	.84	1.715	.99

In the above analyses Group 0 had only 9 *S*s that met the criterion of at least one response in each block of three test trials. Because of possible bias that might enter by such selection of *S*s, a second analysis of magnitude, incorporating data from all *S*s, was made. Table 4 provides the means and *SD*s while Table 5 summarizes the analysis of variance.

Table 5 reveals, first, that there is a significant between-groups effect and, second, that response reliability (as evidenced by the sharply increased *F* value for between-*S*s means) has increased. An examination of mean magnitude in Table 4 indicates that the bulk of the differences between groups rests in Group 0; that is, the increase in *N* in Group 0 over the prior analysis added sufficient numbers of zeros to provide a between-groups effect.

Double responses.—Although preceding analyses were based on the first GSR occurring in the interval from 1 to 5.5 sec. after CS onset, there were, in Groups 3 and 5, frequent instances of a second response with a latency range of from 4 to 8.5 sec. after CS onset. Figure 2, a tracing of one such response from an *S* in Group 5, illustrates the onset of the first response, a plateau, and then the onset of the second response. In Group 5, of 20 *S*s who responded, 19 gave at least one double response. In Group 3, of 18 *S*s who responded, 15 gave at least one double response. With the exception of 1 *S* on one trial in Group 0, no double responses were observed in Groups 0, .5, and 1.

TABLE 5

ANALYSIS OF VARIANCE OF MAGNITUDE FOR ALL *S*s AS A FUNCTION OF ISI AND THREE BLOCKS OF THREE TEST TRIALS EACH

Source	*df*	*MS*	*F*
Between *S*s	105		
ISI	4	7.502	3.71**
Error (b)	101	2.022	9.96***
Within *S*s	212		
Trials	2	2.343	11.54***
ISI × Trials	8	.179	.88
Error (w)	202	2.03	

** $P < .01$.
*** $P < .001$.

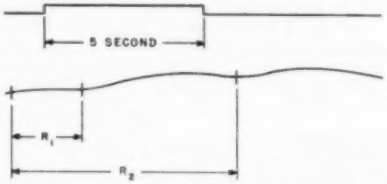

FIG. 2. A tracing of a double response from 1 *S* in Group 5. (Onset of the 5-sec. event marker is temporally coincident with the vertical line indicating initiation of R_1 and R_2, but is displaced due to physical location of event recorder. R_1 and R_2 are latencies of, respectively, the first and second response.)

Discussion

The ISI function obtained with magnitude as the dependent variable is similar to that obtained by White and Schlosberg (1952) and Moeller (1954). Within the range tested, results of all three studies suggest that magnitude is greatest near an ISI of .5 sec., similar to the optimum value obtained in numerous studies of skeletal response conditioning (see Kimble, 1961, pp. 156 f.). That amplitude did not vary as a function of ISI suggests that the magnitude function is determined in large part by whether or not *Ss* in the different ISI conditions responded.[3] On this basis, the distinction between magnitude and amplitude made by Humphreys (1943) and Hilgard (1951) merits further consideration.

Two of the three observations on latency and recruitment have some precedent in the literature: the positive relationship between ISI and recruitment corroborates Bierbaum's (1955) finding and the increase in latency across extinction trials is consistent with Pavlov's (1927, p. 49) report that the latency of salivation increases during extinction trials. We have been unable to find a precedent for the positive relationship between latency and ISI in autonomic conditioning, though such a relationship has been obtained with conditioned skeletal responses (e.g., Boneau, 1958; Ebel, 1961). Though data are limited, present evidence suggests that, in addition to magnitude, the very form of the GSR can be brought under the control of external stimulating events. Such a possibility warrants the continued observation of latency and recruitment in GSR research.

Few references have been made to second responses in GSR studies. Rodnick (1937), with ISIs of 17 and 21 sec., found that a second response would occur shortly before shock onset, and Stewart, Stern, Winokur, and Fredman (1961) demonstrated the acquisition of a second GSR while employing an ISI of 7.5 sec. Grings, Lockhart, and Dameron (1961) have shown, further, that this second response is differentiated more rapidly than is the first response in a discrimination learning situation in which an ISI of 5 sec. is employed. While the nature of the second response is not understood, it tends to occur prior to UCS onset. Whether or not its absence from Groups 0, .5, and 1 results because it does not occur in these groups or because recording speed is too slow to separate response rate changes superimposed on the first response remains to be investigated.

[3] Ratios of number of responders to number in each group on Test Trial 1 were 17/20, 22/22, 16/20, and 17/22 for, respectively, Groups 0, .5, 1, 3, and 5. While there is no entirely adequate statistical test available to compare these frequencies for the five groups simultaneously, a Fisher exact test comparing frequency of responders in Group .5 with that of Group 3 and of Group 5 yielded *P* values of, respectively, .043 and .032 for frequency disparities as large as those obtained. Furthermore, employing the obtained order of mean magnitude (from high to low) in the five groups as a basis for comparison, of the 120 possible ways that the ratios could be ordered only 5 would correspond as well as or better to the magnitude order than that obtained, even if the tied ratio between Groups .5 and 1 were counted as a reversal.

Summary

Five groups of *Ss* trained at CS-UCS intervals of 0, .5, 1, 3, or 5 sec. were employed in a study of the role of the interstimulus interval (ISI) in the conditioning of the GSR. All *Ss* received 20 tone-shock pairings followed by 10 tone alone (test) trials. Four attributes of the GSR were measured: amplitude, magnitude, latency, and recruitment.

The principal findings were: (*a*) magnitude was greatest with an ISI of .5; (*b*) magnitude, but not amplitude, varied with ISI; (*c*) latency and recruitment both increased as a function of ISI; (*d*) latency increased across extinction trials; and (*e*) a second response was observed frequently in *Ss* exposed to ISIs of 3 and 5 sec.

References

Bierbaum, W. B. Temporal aspects in conditioning of the GSR. Unpublished doctoral dissertation, University of Florida, 1955.

BONEAU, C. A. The interstimulus interval and the latancy of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

EBEL, H. C. Stable-state behavior and reversibility as a function of the interstimulus interval variable in classical eyelid conditioning. Unpublished master's thesis, Pennsylvania State University, 1961.

GRINGS, W. W., LOCKHART, R. A., & DAMERON, L. E. Interstimulus interval as a variable in GSR conditioning of mentally deficient individuals. Paper read at American Psychological Association, New York, 1961.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

HUMPHREYS, L. G. Measures of strength of conditioned eyelid responses. *J. gen. Psychol.*, 1943, **29**, 101–111.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning*. (2nd. ed.) New York: Appleton-Century-Crofts, 1961.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

PAVLOV, I. *Conditioned reflexes*. New York: Oxford Univer. Press, 1927.

RODNICK, E. H. Characteristics of delayed and trace-conditioned responses. *J. exp. Psychol.*, 1937, **20**, 409–425.

STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68**, 60–67.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

(Received October 30, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 519–525

ON-TARGET VERSUS OFF-TARGET INFORMATION AND THE ACQUISITION OF TRACKING SKILL[1]

ALTON C. WILLIAMS[2] AND GEORGE E. BRIGGS

Ohio State University

Knowledge of results or information feedback has been an area of substantial research interest since the classic study of Thorndike (1927), the most recent review for motor skills being that of Bilodeau and Bilodeau (1961). The present study was concerned with a restricted topic within the more general area of knowledge of results: the influence of *augmented feedback* on the acquisition of skilled performance. As the term implies, augmented feedback represents knowledge of performance in addition to that normally present in a skill task and it is defined as secondary information supplemental to some primary feedback signal(s).

Of the numerous studies on acquisition of tracking skill only those by Rapparlie (see Bray, 1948, pp. 195–196) and by Payne and Hauty (1955) have employed augmented feedback based on an off-target criterion, the majority of *Es* having activated such signals when tracking error was within tolerance, i.e., when *S* was on-target. Thus, augmented feedback has been utilized in a majority of the research as a means of emphasizing "correct" responding, a use which seems logical in view of the Skinnerian position that errors or incorrect responses should be minimized in the construction of training tasks and programs, at least for discrete-verbal skills. However, it is an open question whether in continuous motor skill tasks *error* information feedback might not be more effective than information on correct responding, and the present study was concerned, therefore, with the effects on tracking performance of augmented feedback based on an off-target criterion relative to that based on an on-target criterion. Further, a simple off-target criterion condition and an off-target condition providing directional information were compared.

METHOD

Apparatus.—The skill task was defined by the SETA apparatus (Gain & Fitts, 1959) which provides *S* with a simple positional tracking task. The *S* tracked a 6 cpm + 12 cpm sinusoidal signal via a one-dimensional compensatory display of tracking information on a 5-in. cathode ray tube. This display was noise free and provided *S* with his primary source of feedback information. Augmented feedback was provided to the experimental groups in the form of auditory clicks delivered to headphones worn by *S*. The clicks occurred at the rate of two per sec. when *S* was within or outside (depending upon the experimental condition) preset tolerance limits of system error. The tolerance limits were attained by adjusting two voltage comparators to the desired voltage levels such that a plate voltage relay was activated whenever tracking error fell within the critical levels. Activation of the relay either closed a circuit from the click generator to the headphones or opened that circuit depending upon the experimental condition. Further, it was possible to activate only the left earphone when tracking error was to the left on

[1] This research was carried out in the Laboratory of Aviation Psychology and was supported by the United States Navy under Contract No. N61339-836, sponsored by the United States Naval Training Device Center, Port Washington, New York. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part for any purpose of the United States Government.

[2] Now with the Space Systems Division, United States Air Force, Los Angeles, California.

the visual display and only the right earphone in the opposite condition. The on-target band thus defined represented tracking error of rather small amplitudes: in terms of the visual display, the on-target band was ±.08 in. around the fixed target element.

Performance was scored by electronically integrating the absolute value of tracking error. This is the analog of the average deviation of *S*'s error amplitude distribution and will be identified as average error. The scores were transformed from voltage units to units of inches on the display scale; thus, average error as reported here may be interpreted as the average deviation of *S*'s tracking error amplitude distribution plotted on the same scale used in the visual tracking display.

In addition to average error, *E* recorded the amount of time during each tracking trial when tracking error was within the tolerance limits. These data were transformed to percent time within tolerance scores.

Subjects and procedure.—The *S*s were 88 volunteer male undergraduates. Each *S* participated in five daily 30-min. sessions, and none of the *S*s had previous experience with a laboratory tracking task.

Twenty-two *S*s were assigned via a randomization procedure to each of four groups. The control group, Group C, did not experience augmented feedback at any time during the experiment. The three experimental groups experienced the auditory clicks during the initial three and one-half training sessions whereupon augmented feedback was withdrawn for the remaining one and one-half sessions. Group I received augmented feedback whenever tracking error was *within* the preset tolerance limits (see above), while Groups O and O-D received the signals whenever tracking error exceeded the on-target limits. For Groups I and O augmented feedback occurred simultaneously in both earphones, while for Group O-D the left earphone was activated when displayed error was to the left of the fixed target reference on the visual display, and the right earphone was activated when error was to the right.

All tracking trials were of 30-sec. duration and were administered in blocks of four trials. There were 3 blocks of trials on the first and last sessions and 4 blocks in each of the other sessions, a total of 18 blocks. The initial 13 blocks defined the training conditions for the experimental groups, while transfer to the no-augmented-feedback condition occurred over the final 5 blocks of trials. Rest periods of 30 sec. were provided between trials within a block and 1.5 min. rest occurred between blocks. Performance was scored over the final 25 sec. of each trial, which permitted the initial transients in tracking behavior to dampen out prior to scoring.

The instructions to all *S*s stressed the importance of maintaining zero error on the visual display, and each *S* was reminded of this ultimate goal at the beginning of each session.

Results

Average error.—Figure 1 provides a summary of tracking accuracy as defined by the average error scores. It may be noted that all three experimental groups attained proficiency levels superior to that of Group C. Further, Group O was superior to both Groups I and O-D throughout both training and transfer, and there was considerable overlap between the latter two groups during training. It appears, then, that augmented feedback based on a simple off-target criterion (Group O) resulted in performance superior to that attained either with an on-target criterion (Group I) or with an off-target criterion which included directional information (Group O-D).

These observations are supported by the results of the Mann-Whitney *U* test which was applied to the average error data for groups. The tracking error scores were summed for each *S* across all 18 trial blocks for this analysis; thus, $n_1 = n_2 = 22$ for each two-group comparison performed. The alternative hypothesis tested here was that more *S*s in one of a pair of groups were superior in tracking proficiency than could occur by chance. A nonparametric analysis was employed since marked heterogeneity of variance was found in the data (see Fig. 2). The results of the *U* tests are listed in the upper half of Table 1, where it may be noted that all group comparisons accepted the alternative hypothesis at $P < .05$ except that between Groups I and O-D.

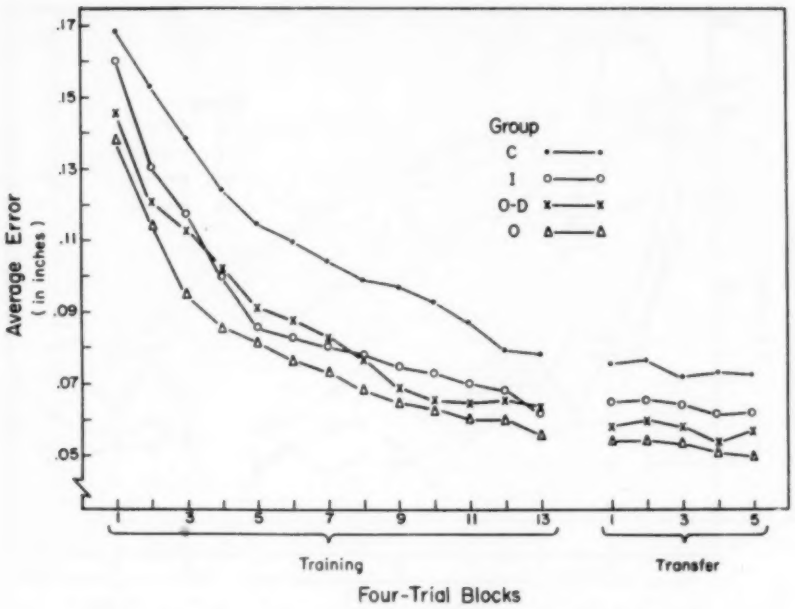

FIG. 1. Average tracking error for all groups during training and transfer.

Variability.—Augmented feedback also exerted an influence on inter-*S* variability within groups (intragroup variability), and these data are summarized in Fig. 2. The data plotted are intragroup *SD*s of the average error scores for each block of training and transfer trials. As with the average error data (Fig. 1), the three experimental groups are superior to the control group in terms of intragroup variability. However, here, Group O does not enjoy a clear superiority over Group O-D, as was the case in Fig. 1, but both Groups O and O-D do exhibit more inter-*S* homogeneity than that attained by Group I.

These observations are supported by the results of *U* tests which were applied to each pairing of groups. It was not possible to test the same hypothesis as in the case of the average error data; instead, the alternative hypothesis tested with the intragroup *SD*s was that more intragroup *SD*s for one of a pair of groups would be smaller in value than those for the other group than could occur by chance. For each comparison $n_1 = n_2 = 18$. The results are listed in the lower half of Table 1 where it may be seen that (*a*) Groups O and O-D exhibited more intragroup homogeneity than did Group I, and (*b*) all three experimental groups were superior in this regard to Group C. It may be concluded, then, that augmented feedback not only results in a higher level of performance accuracy but also individual differences within groups are less than is the case for *S*s who do not experience these additional feedback cues.

Time within tolerance.—The percentage of time *S* spent within the ±.08-in. tolerance limits (TWT) is

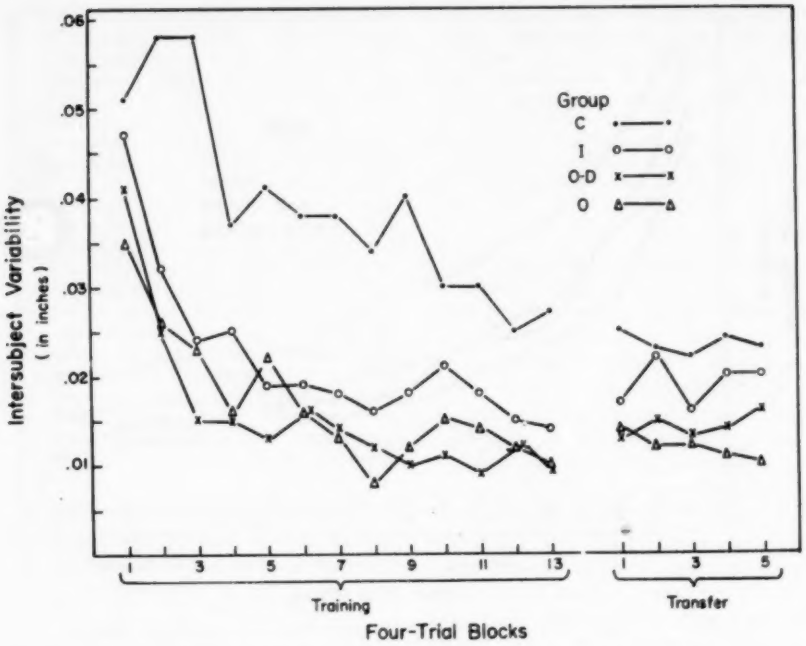

FIG. 2. Inter-*S* variability for all groups during training and transfer.

summarized for each four-trial block in Fig. 3. While the differences among experimental groups are not as apparent as with the average error data, there are statistically significant ($P < .01$) differences both during training and transfer (see Table 2). The pattern of group proficiency levels on TWT differs in an important way from that found with the average error data: in terms of the TWT scores Group I is significantly superior to all other groups during training at $P < .05$. This rank order for Groups

TABLE 1

RESULTS OF THE MANN–WHITNEY *U* TEST APPLIED TO AVERAGE ERROR AHD INTER–*S* VARIABILITY DATA

Performance Criterion		Groups Compared *					
		C/I	C/O-D	C/O	I/O-D	I/O	O-D/O
Average Error ($n_1 = n_2 = 22$)	*U*	333	346	394	270	330	339
	z	2.14	2.44	3.57	0.66	2.07	2.28
	P	.033	.014	.001	.510	.038	.022
Variability ($n_1 = n_2 = 18$)	*U*	33	19	16	46	63	149
	P	<.002	<.002	<.002	<.002	<.002	>.10

* For average error: less/more accurate group; for variability: more/less variable group.

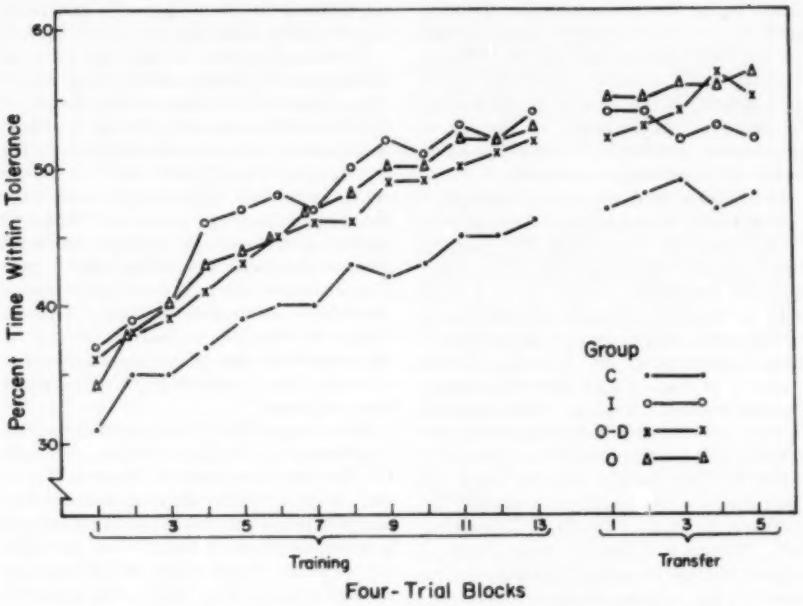

FIG. 3. Percent time within tolerance for all groups during training and transfer.

I and O is just the reverse of that noted for the average error measure where Group O was superior, and this reversal serves to define how *Ss* of Groups O and I responded differentially to the task (see below).

Finally, it may be noted that Group I actually deteriorated during transfer while Groups C, O, and O-D either "held their own" or improved in terms of TWT. This deterioration gave rise to the significant Blocks × Groups interaction in the transfer analysis of Table 2.

TABLE 2

ANALYSIS OF VARIANCE OF PERCENT TIME WITHIN TOLERANCE DATA FOR TRAINING AND TRANSFER

Source	*df*	*MS*	*F*
Training			
Groups (G)	3	2741.13	7.22**
Ss within groups (*Ss*/G)	84	379.54	
Blocks (B)	12	2470.55	182.66***
B × G	36	20.02	1.48*
B × *Ss*/G	1008	13.52	
Transfer			
Groups (G)	3	1314.14	5.34**
Ss within groups (*Ss*/G)	84	246.16	
Blocks (B)	4	17.18	1.21
B × G	12	39.41	2.78**
B × *Ss*/G	336	14.20	

* $P < .05$.
** $P < .01$.
*** $P < .001$.

DISCUSSION

The majority of past research in this area has found superior performance by groups trained with augmented feedback compared to no-augmented-feedback control groups. In this regard, the present research merely confirms those previous data. The contribution of the present study is twofold: first, the use of an off-target criterion for the activation of augmented feedback was found to be superior to an on-target criterion, the

latter being the most common criterion used in previous research; and second, for the first time it has been noted here that not only accuracy but also group homogeneity of tracking performance can be improved significantly by the use of augmented feedback, especially when based on an off-target criterion.

In addition to these major findings, it is of interest to note several other points. First, from Fig. 1 and Fig. 3 it may be seen that no deterioration in performance occurred for either Group O or Group O-D at transfer (augmented feedback withdrawn), but Group I does exhibit some deterioration on Transfer Blocks 1 and 2 of Fig. 1 and over the entire transfer session in Fig. 3. This is logical in view of the similarity between the training and transfer trials for Groups O and O-D, i.e., during training these *S*s experienced a diminution in number of auditory clicks as a result of increasing skill. Group I, however, experienced a marked change in going from training to transfer, i.e., these *S*s experienced increasing amounts of auditory clicks with training followed by an abrupt change to no-augmented-feedback at transfer. These observations were supported by statistical analyses of the transfer data: Groups O and O-D were superior to Group C throughout transfer in terms of tracking accuracy ($P < .05$), but Groups I and C did not differ ($P > .05$) from the first transfer trial to the last.

The reason for the loss of superiority by Group I over Group C may be found in Fig. 2: upon transfer Group I exhibited an increase in inter-*S* variability considerably greater than the modest increase shown by Groups O and O-D. It follows, then, that the transfer performance deterioration by Group I involved not only a loss in tracking accuracy (Fig. 1) but also an increase of within-group variability (Fig. 2). Since apparent accuracy deterioration is not noted for Groups O and O-D, and since the increase in inter-*S* variability was relatively less than that for Group I during transfer, it follows that an off-target criterion for the activation of augmented feedback is to be preferred as a training variable.

A second point of interest can be introduced by again noting from Fig. 1 that, in terms of average error, Group O is superior throughout to Group I. Now, the average error metric actually is the average deviation (AD) of *S*'s tracking error amplitude distribution, and thus the smaller average error for Group O indicates that, on the average, those *S*s generated smaller errors than did Group I. However, Group I actually spent more time within the ±.08-in. tolerance limits (see Fig. 3). It follows, then, that the *shape* of the error amplitude distribution for Group O must differ from that of Group I.

What might be the forms of the error amplitude distributions for Groups O and I? The previous data of Bahrick, Fitts, and Briggs (1957) suggest that if the distribution of tracking error for Group I is nonnormal (as in fact is the case for the present data), that distribution is probably leptokurtic. On this basis, it follows that Group O generated a *bimodal* error amplitude distribution, as this is the only distribution that could result in both smaller AD and smaller percent time within tolerance scores for Group O compared to Group I. In other words, *S*s of Group O spent considerable time tracking closely *around* the two tolerance limit points and made relatively few errors of large amplitude, while Group I spent more time *within* the tolerance limits but committed occasional large tracking errors.

These deductions from the data, then, lead one to conclude that tracking behavior with augmented feedback based on an off-target criterion differs fundamentally from that when such feedback is based on an on-target criterion: in the former case large errors are emphasized, and *S* apparently learns rather quickly to minimize these occasional lapses in tracking accuracy, while in the latter case the importance of small errors is emphasized, but apparently *S* does not respond as quickly or as efficiently to correct for occasional large tracking errors. Since minimizing large tracking

errors is a primary task for *S* early in training, it follows that augmented feedback based on an off-target criterion should be particularly helpful in shaping the desired behavior during early acquisition of a continuous control skill. It is problematical whether or not the above generalization will hold in discrete-verbal learning tasks. However, it is suggested that if the element of discovery is present in such discrete tasks, some criterion for the activation of augmented feedback analogous to the off-target criterion in the continuous case would be helpful in focusing *S*'s attention on response alternatives reasonably close to the correct alternative.

Finally, it is interesting to note that the additional feedback information available to Group O-D (off-target direction of error) did not provide for a performance level superior to that of Group O which received augmented feedback on a more simple, nondirectional off-target criterion.

Two explanations for the relative inferiority of Group O-D are suggested. First, it is possible, of course, that the addition of a directional cue to the off-target criterion was not particularly useful information to Group O-D since directional relationships between control and display movements are one of the most simple aspects of a tracking task to be learned. Secondly, it is possible that the information on error direction was actually disruptive. It may be recalled that auditory clicks were delivered to the left earphone when error was to the left on the visual tracking display. Thus, a signal in the *left* earphone indicated that *S* should move his control *to the right*. While data are lacking on population stereotypes in such a stimulus-response task, it is probable that a more compatible S-R arrangement would be one in which a signal in, say, the *left* earphone indicates a control movement is required *to the left*. It is suggested that both of the above possibilities were responsible for the inferior performance of Group O-D relative to that of Group O.

Summary

During training three experimental groups received augmented feedback (auditory clicks at the rate of two per sec.) when tracking accuracy was within (an on-target criterion) or outside (an off-target criterion) fixed tolerance limits. During transfer, no augmented feedback was provided. All three experimental groups were superior in tracking accuracy during training to a control group which did not receive augmented feedback. Of the three experimental groups, the group receiving augmented feedback when off-target was superior to a group which experienced clicks when on-target. It was superior also to a group which received clicks when off-target but differentially according to the direction of tracking error.

During transfer both off-target groups remained superior in tracking accuracy to the control group, but the on-target group and the control group attained comparable performance.

It follows that augmented feedback based on a simple off-target criterion was the most effective training condition. An analysis of the data suggested that this superiority was a result of the emphasis an off-target criterion places on occasional large tracking errors. The group trained on this condition apparently learned to reduce such errors more quickly and efficiently than did the on-target criterion group.

References

Bahrick, H. P., Fitts, P. M., & Briggs, G. E. Learning curves: Facts or artifacts? *Psychol. Bull.*, 1957, **54**, 256–268.

Bilodeau, E. A., & Bilodeau, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

Bray, C. W. *Psychology and military proficiency.* Princeton, N. J.: Princeton Univer. Press, 1948.

Gain, P., & Fitts, P. M. A simplified electronic tracking apparatus (SETA). *USAF WADC tech. Rep.*, 1959, No. 59-44.

Payne, R. B., & Hauty, G. T. Effect of psychological feedback upon work decrement. *J. exp. Psychol.*, 1955, **50**, 343–351.

Thorndike, E. L. The law of effect. *Amer. J. Psychol.*, 1927, **39**, 212–222.

(Received October 30, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 526–532

RESISTANCE TO EXTINCTION AFTER VARYING AMOUNTS OF DISCRIMINATIVE OR NONDISCRIMINATIVE INSTRUMENTAL TRAINING[1]

M. R. D'AMATO, DONALD SCHIFF, AND HARRY JAGODA

New York University

Resistance to extinction and number of reinforced responses (acquisition level) have classically been thought to be monotonically related. Recently, however, doubt concerning the monotonicity of the relationship has been raised in several quarters (e.g., Birch, Ison, & Sperling, 1960; Murillo & Capaldi, 1961; Senko, Champ, & Capaldi, 1961). Although nonmonotonicity between resistance to extinction and acquisition level has been reported in a rather large number of recent studies, many of these are of questionable relevance because either they made use of some type of an intermittent reinforcement schedule (e.g., Capaldi, 1957, 1958) or they employed a discriminative rather than a nondiscriminative task (Murillo & Capaldi, 1961), in some cases "extinction" actually constituting reversal learning (e.g., Senko et al., 1961). In certain other studies reporting nonmonotonicity and not subject to either of the preceding objections (Lewis & Duncan, 1956, 1958), one finds other reasons for questioning their pertinence for the present problem, such as the use of human *Ss* in money payoff situations, which bear only the slightest resemblance to the strictly instrumental situations in which, with animal *Ss*, the relationship of monotonicity was originally established.

[1] This research was supported by Research Grant M-2051 from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service, and Grant G-14724 from the National Science Foundation.

Of the several published studies falling in the latter category, two have reported resistance to extinction and acquisition level to be nonmonotonically related. North and Stimmel (1960) found that extinction of the running response proceeded more rapidly for *Ss* given 90 or 135 rewarded trials in a runway than for *Ss* given 45 reinforced trials. Wilson (1958), on the other hand, found no evidence of a reduction in resistance to extinction, even though he carried (runway) acquisition to 480 trials. Harris and Nygaard (1961), working with thirsty rats in a free operant situation, observed the usual monotonic relationship between acquisition level and resistance to extinction of the bar pressing response up to 360 reinforced responses, the highest acquisition level they worked with. Under rather similar conditions, Margulies (1961) reported monotonicity up to 1,000 trials; and as described in an earlier paper (D'Amato & Jagoda, 1962), unpublished results from our laboratory suggest persistence of the monotonic relationship up to some 7,000 rewarded bar pressing responses, again with thirsty rats as *Ss*. Finally, in conflict with the preceding results, King, Wood, and Butcher (1961) recently reported a reduction in the resistance to extinction of pigeons receiving 600 or 900 reinforced key pecks, as compared to *Ss* permitted 300 rewarded responses.

In summary, the results from studies employing a nondiscriminative (simple) instrumental response are equivocal, though they suggest that under certain conditions the monotonicity assumption holds up to several thousands of responses.

The present study had three objectives. First, all of the cited Skinner box studies in which monotonicity was obtained employed the thirst drive and water reward. It would be of some value to demonstrate that monotonicity also holds in the Skinner box

setting for the hunger drive and food reward, particularly since all other relevant studies have been conducted under the latter motive-incentive conditions. Second, we wished to carry acquisition training beyond the level achieved in previously published reports, to a maximum of 1,600 reinforcements. Third, and most important, we wished to assess the influence of two different training procedures, discriminative versus nondiscriminative instrumental training, on the shape of the function relating acquisition level and resistance to extinction. The hypothesis under examination was that the traditional monotonic relationship would prevail for *S*s given nondiscriminative training, but not for *S*s given discrimination training, the latter being expected to show a significant decline in resistance to extinction with prolonged acquisition training.

One plausible rationale for the preceding hypothesis is as follows. In general, *S*s trained on a discrimination program are essentially on an intermittent reinforcement schedule (with respect to the experimental situation) until the discrimination is firmly acquired; beyond this point their experience is to all intents and purposes one of continuous reinforcement (since they then make few errors, or respond little in S^{Δ}). Prolonged discrimination training, then, has the effect of carrying *S*s beyond their intermittent reinforcement experience well into the continuous reinforcement segment. If it is assumed that directly following an intermittent reinforcement experience resistance to extinction is greater than when a prolonged continuous reinforcement segment intervenes between the intermittent reinforcement experience and extinction, the hypothesis follows.

Method

Subjects.—The *S*s that completed the study were 94 experimentally naive albino rats (51 males and 43 females) 53 to 94 days of age at the start of the study. One *S* was discarded because of a programing error and a second due to illness. All *S*s were bred in our laboratory.

Apparatus.—Two Grason-Stadler two-bar Skinner boxes were used; the right bars were removed from both boxes, converting them into single-bar boxes. A force of approximately 20 gm. was required to activate the bar microswitches in the two boxes.

Design.—The experiment was run in two replications (of unequal *N*), but since the results of the two replications were quite similar in form, replications was not included as a factor in the statistical design.

There were four levels of acquisition, defined in terms of the number of reinforced responses allowed after pretraining: 200, 400, 800, and 1,600. At each acquisition level there were two separate groups of *S*s that differed in the type of training received. The "I" groups, designated as I(200), I(400), I(800), and I(1600), underwent simple (nondiscriminative) instrumental training, each of the four groups being composed of 12 *S*s. The "D" groups, D(200), D(400), D(800), and D(1600), were trained on a successive brightness discrimination problem. There were 11 *S*s in the first two groups and 12 each in the last two groups.

All *S*s of a replication were quasirandomly assigned to the eight groups, in part balancing litters over groups, and placed on deprivation at the same time. The I and D groups of a given acquisition level started acquisition training together; the scheduling of the beginning of acquisition was so arranged for the various reinforcement groups that all groups entered extinction within a day or two of each other.

Deprivation training.—One week prior to the beginning of pretraining, *S*s were placed on a 22-hr. food deprivation training regimen, water being constantly available in the home cages. Four days later, the feeding period was reduced to 1½ hr. daily and remained at that duration throughout the study.

Skinner box pretraining.—There were 3 pretraining days. On Days 1 and 2, 50 reinforcements (45-mg. Noyes rat pellets) were given in the conditioning of approach responses to the food tray at the sound of the feeder magazine. On Day 3, *S*s were shaped on the bar pressing response with 25 to 50 reinforcements. The stimulus conditions in

the boxes during pretraining were the same as prevailed during simple instrumental training and during the S^D portion of discrimination training.

Discrimination training.—All D groups were exposed to a simple brightness discrimination situation. Illumination of the left (white jeweled) stimulus light constituted S^D and provided an illumination level of 7 to 10 ft-c, measured with the target of a Weston illumination meter placed at bar level and ¾ in. in front of the lens of the stimulus light. Measured under the same conditions the illumination under S^Δ, which was produced by the shielded house light, read about 0.1 ft-c.

All S^D periods were 45 sec. in duration and alternated with S^Δ periods varying between 33 and 65 sec. in length. One revolution of the programing film tape took about 6 min. and provided approximately equal total times in S^D and S^Δ. Since 50 reinforced responses in S^D were allowed each day, Groups D(200), D(400), D(800), and D(1600) required 4, 8, 16, and 32 training days, respectively.

Simple instrumental training.—The treatment of the I groups was exactly the same as that accorded the corresponding D groups except for the elimination of the S^Δ periods, i.e., the left stimulus light was always on. As in the D groups, 50 reinforcements were permitted on each acquisition day.

At the end of the daily 50 reinforced responses, *Ss* of all groups entered a time-out period and were quickly removed from the Skinner boxes to their home cages where, no sooner than 15 min. later, they were fed for the 1½-hr. period. The estimated average number of hours of food deprivation at the start of a day's session was 22 hr.

Extinction.—The extinction procedure, which began the day following the termination of acquisition, was precisely the same for all *Ss* and consisted of one daily 10-min. period on each of 5 successive days with the left stimulus light illuminated, i.e., in the former S^D condition. The duration of the extinction periods, 10 min., was judged to be short enough to resemble closely the length of the acquisition sessions (which overall averaged 7.15 min.) and yet long enough to sample adequately the extinction process. It should be pointed out, however, that because of the absence of S^Δ periods, the extinction sessions bore a greater similarity to the acquisition sessions of simple instrumental training than to those of discrimination training. The exclusion from extinction of S^Δ periods (more correctly, periods in which the former S^Δ stimulus was present) was based on two considerations: (*a*) We wished to make direct comparisons of the extinction performance of corresponding I and D groups, which would be feasible only if the extinction sessions were identical for all *Ss*. (*b*) Inclusion of such periods would have raised problems concerning the treatment of responses made in S^Δ, since the number of such responses most probably would be related to the acquisition level variable.

It should be recorded that the cue provided by activation of the feeder magazine was maintained during extinction. Deprivation level at the start of extinction sessions was confined within the limits of 21 to 23 hr. During the course of the experiment the relative humidity ranged between 58% and 70%, and the temperature between 70° and 78° F.

Results

Discrimination acquisition.—A discrimination ratio (DR), obtained by dividing the number of responses made in S^Δ by the number performed in S^D, was calculated for each *S* after every daily training session. In a plot of the daily means of the DRs, Group D(1600) showed some improvement over the level of discrimination finally achieved by Group D(800). To evaluate this difference the mean DR over the last 8 acquisition days was calculated for every *S* of Group D(1600), as was the mean DR over the last two sessions for each *S* of Group D(800). The group means based on these measures (.141 and .218, respectively) did not, however, differ significantly ($t = 1.67$, $df = 22$, $P \approx .12$).

Extinction in the I groups.—The number of responses made by each *S* in each of the five daily extinction sessions was converted to common logs, providing the basis for all statistical analyses. An overall index of resistance to extinction was obtained for each *S* by summing its five daily log scores.

The first question of interest concerns the relationship observed be-

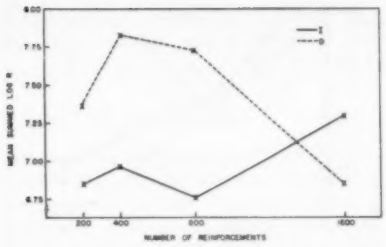

FIG. 1. Relationship between acquisition level (number of reinforcements) and resistance to extinction in the simple instrumental (I) and the discrimination (D) groups.

tween resistance to extinction and the number of reinforced responses allowed during acquisition. The solid line in Fig. 1, based on the means of the summed log scores, presents this relationship. It is plain from the figure that there is no tendency for resistance to extinction to decrease with increasing training, even when acquisition is carried to 1,600 reinforced responses; in fact, the solid line suggests an increasing trend.

The most powerful (and specific) way of analyzing the present data is by a trend analysis of the linear and quadratic components of the curve (Grant, 1956). The presence of a significant quadratic component (with negative sign) alone or in accompaniment with a significant negative linear component would constitute evidence for a nonmonotonic relationship between acquisition level and resistance to extinction. A significant negative linear component in conjunction with an insignificant quadratic would, essentially, support the same interpretation. On the other hand, an insignificant quadratic component would support the assumption of monotonicity if the linear component were either insignificant or positive in sign. The latter would indicate that the curve was still rising, while the occurrence of insignificance in both the linear and quadratic components would signify that the function was asymptotic.

In order to take into consideration the unequal spacing of the independent variable, the coefficients of the orthogonal polynomials were calculated in the manner suggested by Grandage (1958). Analysis of variance showed both the linear and the quadratic components of the trend to be insignificant ($F = 1.29$ and 0.59, respectively, $df = 1/44$). Thus, for the numbers of rewarded responses employed in this study, the function relating resistance to extinction and acquisition level in the nondiscriminatively trained *Ss* was essentially asymptotic.

Despite the absence of significant differences in the summed log extinction scores, one would like to know whether extinction in the four I groups followed a parallel development over the five extinction sessions. Figure 2 presents the extinction curves of the four I groups over the 5 extinction days. A trend analysis

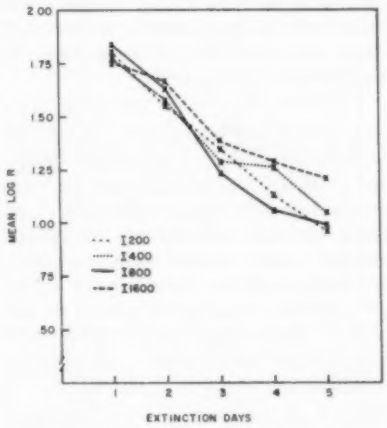

FIG. 2. Extinction curves of the four simple instrumental groups, based on the log of the number of daily extinction responses.

(Edwards, 1960) was applied to the data and the differences among the groups' linear components evaluated. The appropriate F (2.38, $df = 3/176$) fell short of accepted significance levels ($P \approx .07$). The curves of Fig. 2 suggest that more impressive differences among the linear components of the groups' trends, as well as among their overall extinction scores, might have been obtained if extinction had been carried one or two sessions further. At any rate, the data of the figure plainly show that Group I (1600) is not inferior to any of the other I groups.

Extinction in the D groups.—The first concern is again with the relationship between acquisition level and resistance to extinction, which, as may be seen from Fig. 1, is vastly different from that obtained with the I groups. A trend analysis of the curve revealed a significant linear component ($F = 4.13$, $df = 1/42$, $P < .05$), as well as a significant quadratic component ($F = 6.31$, $df = 1/42$, $P < .025$). Because of the unequal Ns in the four discrimination groups, the trend analysis was based on groups means, rather than sums, in the manner suggested by Walker and Lev (1953).

Figure 3 shows that the inferiority of Group D (1600) relative to Groups D (400) and D (800) is present throughout extinction, though it appears most marked on the last 2 extinction days. Once again the differences among the groups' linear trends were evaluated and once again the resulting F (2.51, $df = 3/168$) was quite close to accepted significance levels ($P \approx .06$), suggesting differences among the slopes of the extinction curves.

Thus, in contrast to the I groups, the discriminatively trained Ss, after an initial increase, reveal a sharp reduction in resistance to extinction as a function of increasing acquisition training. Further, the nonmonotonicity appears in the early stages of extinction, tending to be somewhat more marked during the latter extinction days.

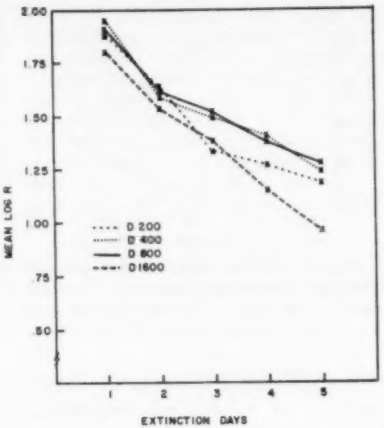

FIG. 3. Extinction curves of the four discrimination groups, based on the log of the number of daily extinction responses.

Comparison of extinction in the I and D groups.—If it is in fact true that discrimination training, at least in its early phases, provides an intermittent reinforcement experience, then resistance to extinction should be greater in the combined D groups than in the I groups. Analysis of the differences between the means of the summed log scores of the combined four D groups (7.43) and the four I groups (6.96) revealed that discrimination training did indeed lead to significantly greater resistance to extinction ($t = 2.53$, $df = 86$, $P < .02$).

It will be observed in Fig. 1 that a reversal in the extinction curves of the I and D groups occurs in the 1600 groups; the reversal is, however, far from significant ($t = 1.25$, $P > .20$). Finally, tests of differences between the means of the I and D groups receiving 200, 400, and 800 reinforce-

ments yielded, in the same order, the following results: $t = 1.45$, $P > .10$; $t = 2.25$, $P < .05$; $t = 2.63$, $P < .02$.

Discussion

In general, our results support the hypothesis that the presence or absence of monotonicity in the function relating acquisition level to resistance to extinction depends importantly on the type of learned response under consideration. In agreement with earlier reports (Harris & Nygaard, 1961; Margulies, 1961) monotonicity was found with a nondiscriminative (free operant) response; with a discriminative response, on the other hand, a marked nonmonotonicity between acquisition level and resistance to extinction was observed. While the present results do little to explain the nonmonotonicity obtained in the presumably nondiscriminative situations of North and Stimmel (1960) and King, Wood, and Butcher (1961), they perhaps shed some light on other, related, studies in which a strong discriminative component was present (e.g., Murillo & Capaldi, 1961; Senko et al., 1961). Conceivably, they also possess some relevance for the overlearning reversal effect (the faster reversal learning of *Ss* receiving extensive overtraining), inasmuch as that phenomenon has been attributed by some to a nonmonotonic relationship between acquisition level and resistance to extinction of the approach response to the originally positive stimulus (e.g., Birch et al., 1960).

Turning now to a possible mechanism by which the nonmonotonicity of the discrimination groups might be explained, two separate factors seem to be involved. First, it probably is safe to assume that the intermittent reinforcement experience unavoidably associated with the early phases of discrimination training has the effect of augmenting resistance to extinction in *Ss* receiving moderate amounts of discrimination training. This assumption is supported in the present study by the superior resistance to extinction of the combined discrimination groups, with further verification coming from an earlier study by Jenkins (1961a), who worked with pigeons in a discrete trials situation.

Second, it is still necessary to explain how resistance to extinction becomes depressed with overtraining, and two possibilities suggest themselves. The first one is that mentioned earlier, namely, that overtraining has the effect of carrying *Ss* beyond their intermittent reinforcement experience well into a region of virtual continuous reinforcement. However, there are the following difficulties with this possibility. (*a*) Several studies (e.g., Jenkins, 1961b; Theios, 1962) have failed to demonstrate a reduction in resistance to extinction as a result of interpolating a continuous reinforcement segment between a partial reinforcement experience and subsequent extinction. (*b*) There is evidence that nonmonotonicity between acquisition level and resistance to extinction can occur in situations where responding to the negative stimulus is under the control of *E* rather than *S*, as in successive discrimination training on a straightaway (Birch et al., 1960). (*c*) Finally, the nature of the present argument is such that, in principle, no amount of discrimination training could reduce resistance to extinction below that of a comparable nondiscriminative group. The results obtained with the 1,600 groups of the present study suggest that such a reversal is a distinct possibility if acquisition were carried somewhat further.

An alternative interpretation of the basis of the nonmonotonicity maintains that the discrimination experience is vital completely apart from attending changes in the effective reinforcement schedules (cf. Murillo & Capaldi, 1961, who found nonmonotonicity only in those *Ss* that had learned the discrimination presented them during acquisition). This is a position, however, that is difficult to specify with any degree of precision. Nevertheless, it is suggestive that in Group D(1600) goodness of discrimination (the reciprocal of the mean of the DRs over the last 10 acquisition days) was negatively cor-

related with resistance to extinction (summed log R), with rho equal to .53 ($.05 < P < .10$).

Summary

This experiment investigated the hypothesis that acquisition level and resistance to extinction would be monotonically related for a simple (nondiscriminative) instrumental response (bar pressing in a Skinner box), but the function would be nonmonotonic for a comparable discriminative response (successive brightness discrimination). Four groups of *Ss* were trained on the simple instrumental response and allowed 200, 400, 800, or 1,600 reinforced responses, 50 per day. The same numbers of reinforced responses were given to four corresponding groups of discriminatively trained *Ss*, the procedure employed with the latter differing only in the insertion of occasional S^{Δ} periods. All groups were exposed to one 10-min. extinction period (in the former S^{Δ}) on each of 5 successive days. Trend analyses of the data supported the initiating hypothesis; and as expected, the discriminatively trained *Ss* were, as a group, more resistant to extinction than the *Ss* trained on the simple instrumental response.

References

Birch, D., Ison, J. R., & Sperling, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

Capaldi, E. J. The effects of different amounts of alternating partial reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1957, **70**, 451–452.

Capaldi, E. J. The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, **51**, 367–371.

D'Amato, M. R., & Jagoda, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, **64**, 117–122.

Edwards, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Holt, Rinehart, & Winston, 1960.

Grandage, A. Orthogonal coefficients for unequal intervals. *Biometrics*, 1958, **14**, 287–289.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Harris, P., & Nygaard, J. E. Resistance to extinction and number of reinforcements. *Psychol. Rep.*, 1961, **8**, 233–234.

Jenkins, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, **61**, 111–121. (a)

Jenkins, H. M. Resistance to extinction when partial is followed by regular reinforcement. Paper read at Psychonomic Society, New York, 1961. (b)

King, R. A., Wood, P., & Butcher, J. *Decreased* resistance to extinction as a function of reinforcement. *Amer. Psychologist*, 1961, **16**, 468. (Abstract)

Lewis, D. J., & Duncan, C. P. The effect of partial reinforcement and length of acquisition-series upon resistance to extinction of a motor and a verbal response. *Amer. J. Psychol.*, 1956, **69**, 644–646.

Lewis, D. J., & Duncan, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55**, 121–128.

Margulies, S. Response duration in operant level, regular reinforcement, and extinction. *J. exp. Anal. Behav.*, 1961, **4**, 317–321.

Murillo, N. R., & Capaldi, E. J. The role of overlearning trials in determining resistance to extinction. *J. exp. Psychol.*, 1961, **61**, 345–349.

North, A. J., & Stimmel, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

Senko, M. G., Champ, R. A., & Capaldi, E. J. Supplementary report: Resistance to extinction of a verbal response as a function of the number of acquisition trials. *J. exp. Psychol.*, 1961, **61**, 350.

Theios, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 1–6.

Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.

Wilson, J. J. Level of training and goal-box movements as parameters of the intermittent reinforcement effect. Unpublished doctoral dissertation, New York University, 1958.

(Received November 14, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 533–540

T MAZE REVERSAL LEARNING AFTER SEVERAL DIFFERENT OVERTRAINING PROCEDURES[1]

WINFRED F. HILL, NORMAN E. SPEAR
Northwestern University
AND KEITH N. CLAYTON
Vanderbilt University

Several investigators (Brookshire, Warren, & Ball, 1961; Capaldi & Stevenson, 1957; North & Clayton, 1959; Pubols, 1956; Reid, 1953) have demonstrated that overlearning of a discrimination facilitates its subsequent reversal. This effect at first seems paradoxical, since it appears to imply that increasing the number of reinforced trials to one cue weakens the tendency to respond to that cue.

Several mechanisms have been suggested which might explain this overlearning-reversal effect (ORE). (*a*) The additional practice in making choices during the overtraining may facilitate subsequent reversals. This might be mediated by acquired observing responses, as suggested by Reid (1953) and Pubols (1956). (*b*) The long series of rewards may make a change more discriminable and hence make reversal learning faster, a suggestion made by Capaldi and Stevenson (1957). (*c*) The greater number of rewards, by building up stronger r_G's, may result in greater frustration and hence greater disruption when reward ceases to follow the accustomed response. This is the explanation offered by North and Stimmel (1960) for their finding that overlearning facilitated extinction in a straight alley. It is also consistent with Birch, Ison, and Sperling's (1960) finding that with single stimulus presentation the overlearning-reversal effect is mainly attributable to the rate of extinction of the old response. (*d*) The long series of trials on which nearly every response is rewarded may reduce, through any of various mechanisms, the tendency to avoid the incorrect cues, thus making it easier to approach those cues when they become correct. This explanation is suggested by D'Amato and Jagoda (1961). (*e*) So much stimulus satiation may be built up to the correct cues that there is a tendency to avoid these cues as soon as they cease to be associated with reward. Though this has not been suggested as an explanation for the ORE, it seems consistent with research summarized by Glanzer (1958).

The present study attempted to provide evidence concerning the relative importance of these various mechanisms for the ORE. Four groups were compared. Two were the groups found in any study of the ORE: Group N reversed as soon as the discrimination was learned, Group Fr given overtraining. The other two groups were also given overtraining, but with special features. Group Co was forced to the correct side on all its overtraining trials, thus being deprived of the opportunity for making choices. Group In was given twice as many overtraining trials as Groups Fr and Co, with half forced to the correct and half to the in-

[1] This research was supported by Grant G-8706 from the National Science Foundation and was conducted at Northwestern University.

correct side. This group was thus given more experience with the incorrect side than any of the others. It was predicted that Group Fr would reverse fastest and either Group N or In slowest, with the exact positions of the groups throwing light on the various explanations of the ORE.

EXPERIMENT I

Method

Subjects and apparatus.—The *S*s were 64 experimentally naive female albino rats of the Sprague-Dawley strain, 74 to 75 days old at the beginning of experimental training. One *S* was discarded because of apparatus failure and was replaced by another *S* of the same description. The apparatus was the narrow T maze described by Cotton, Lewis, and Metzger (1958). This is an enclosed T maze with a 4-ft. stem and 9-in. arms (excluding goal box). The wooden doors beyond the choice-point, utilized to prevent retracing, were also used for forcing purposes in this experiment.

Prehandling.—Each *S* received six daily 3-min. sessions of prehandling, the last one 48 hr. prior to the beginning of experimental training. During each session *S* was allowed to explore a large unpainted wooden box, presented with four of the pellets later to serve as reward, and picked up and replaced at least five times by *E*. Pellets uneaten were returned with *S* to the home cage along with a fifth pellet. A once-daily feeding schedule began on the first day of prehandling and was maintained throughout experimental training. The ration was 10 gm. of finely ground Purina lab chow and was presented 30–35 min. after the start of prehandling or experimental training.

Experimental training.—On each trial, after placement into the start box, the door was opened as soon as *S* was oriented toward it. Located just outside the start box was a treadle which, when *S* stepped on it, started two 1/100 sec. clocks. The first clock stopped automatically when *S* stepped on a treadle located just before the choice-point. The second clock stopped when *S* stepped on a treadle just beyond the choice-point in either arm of the maze. The side designated as correct for a particular *S* was baited with two 45-mg. Noyes pellets presented in a slightly bent tin lid. No lid was present on the nonrewarded side. For half the *S*s, the right side was designated correct; for the others the left was correct. The noncorrection procedure was used throughout, and the duration of goal-box confinement was approximately 10 sec. regardless of choice. Between trials *S* was kept for 15 sec. in a carrying cage before being placed in the start box for the next trial.

General design.—There were three stages to the experimental training—acquisition, overtraining, and reversal. The experimental variable was manipulated only in the overtraining stage. However, prior to the beginning of acquisition, *S*s were randomly assigned to four experimental groups of 16 *S*s each. The three stages of training required a total of 13 days for all groups.

Acquisition.—Twelve free trials were given on each of 3 consecutive days. Groups Fr, Co, and In received acquisition on Days 1, 2, and 3 of acquisition. Group N, however, received acquisition on Days, 1, 2, and 11.

Overtraining.—On Days 3–6 and 8–10 Group N was treated exactly as it had been during prehandling and was given no T maze experience during this time. It thus served as a no-overtraining control group. Group Fr was given 15 free trials on each of Days 4–6 and 8–10 and 12 free trials on Day 11, making a total of 102 free overtraining trials. Group Co was given 15 trials, all forced to the correct side, on each of Days 4–6 and 8–10, and 12 trials, also forced correct, on Day 11. This made a total of 102 forced-correct overtraining trials. Group In was given 30 trials on each of Days 4–6 and 8–10 and 24 trials on Day 11. All of these 204 trials were forced, half to the correct side and half to the incorrect. The distribution of correct and incorrect forced trials was determined randomly, and was different for each *S*. On Day 7, all *S*s in all groups were fed on schedule but were not handled.

Reversal.—All *S*s received 15 free trials on each of Days 12 and 13. During this stage, reward was placed in the goal box opposite to that which had originally held the reward during acquisition and overtraining, so that the formerly correct goal box became incorrect and vice versa.

Results

Acquisition.—The mean proportion of correct choices on the last five trials was .93. An analysis of variance on the latter half of acquisition revealed no significant differences among the groups ($F = 1.87$). In Group N there was no evidence of forgetting

TABLE 1

TOTAL CORRECT CHOICES IN REVERSAL

Group	Experiment I		Experiment II		Experiment III	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
N	21.88	2.92	20.31	5.69	18.15	7.40
Fr	18.94	1.48	16.69	5.42	9.07	9.13
Co	21.31	2.75	21.38	4.50		
In	10.75	5.23	7.06	6.10		

over the interval between the second and third days of training.

Overtraining.—Group Fr made a total of 3.3% incorrect responses during overtraining. The median number of correct responses after the last error was 54.

Reversal.—The course of reversal learning for the four groups is shown in Fig. 1. The mean number of correct choices on the 30 reversal trials is shown for the four groups in Table 1. The overall F for the four groups was 37.01, significant at the .001 level. Adjacent groups were compared with t tests. Groups N and Co did not differ significantly ($t = .57$). However, Group Co was superior to Group Fr at the .01 level ($t = 2.96$), and Group Fr in turn was superior to Group In at the .001 level ($t = 6.02$).

The mean number of incorrect responses before the first correct reversal response is shown for the four groups in Table 2. All differences were significant at the .01 level except that between Groups N and Co, which did not approach significance.

In view of the failure to find an overlearning-reversal effect in terms of total correct responses in reversal, the reversal data were also analyzed in terms of a criterion of 18 correct choices in 20 successive trials, with not more than one error in the last 10 of the 20. This is the same as Pubol's (1956) criterion except that his 20-trial units always involved 2 complete days of running, whereas ours could begin on any trial. Comparison with Pubol's second experiment is most appropriate since it was closest to ours in the kind of discrimination involved. The numbers of *S*s reaching this criterion in the four groups were 10, 7, 10, and 1, respectively. These values yield a χ^2 of 13.7, significant at the .01 level for 3 *df*.

Speeds.—Speeds in feet per second on the first five trials of reversal are shown in Table 3. Stem speeds are from the starting treadle to the treadle just before the choice-point; total speeds are from the starting treadle to one of the two treadles just beyond the choice-point. For stem speeds there were no significant differences among the groups. The overall F was 1.29, and the t ratio between the two extreme groups (1 and 2) was 1.79. For total speeds, however, the overall F was 4.16, significant at the

TABLE 2

NUMBER OF ERRORS BEFORE FIRST CORRECT RESPONSE IN REVERSAL

Group	Experiment I		Experiment II		Experiment III	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
N	3.19	1.83	6.44	5.75	5.08	3.35
Fr	6.19	1.17	9.31	5.93	11.93	8.50
Co	3.81	3.02	5.25	3.77		
In	12.44	5.45	16.19	8.90		

TABLE 3

MEAN SPEEDS IN FEET PER SECOND ON FIRST FIVE REVERSAL TRIALS

Group	Experiment I		Experiment II		Experiment III	
	Stem	Total	Stem	Total	Stem	Total
N	4.01	2.70	3.87	2.89	3.91	2.86
Fr	4.42	3.50	4.36	3.53	4.20	3.23
Co	4.11	3.03	4.45	3.48		
In	4.07	3.10	3.44	2.78		

.01 level. Of the differences in total speed among adjacent groups, only that between Groups Fr and In reached the .05 level of significance by *t* test.

Experiment II

The failure to find an ORE in Exp. I brought the purpose of the study into question. An attempt was therefore made to change the conditions so as to replicate the earlier findings of an ORE. In order to bring Exp. II closer to these earlier studies, the number of overlearning trials was increased, the intertrial interval was made longer, and various other details were changed.

Method

Subjects and apparatus.—The *S*s were 64 experimentally naive female albino rats of the Sprague-Dawley strain, 74 to 78 days old at the beginning of experimental training, divided into four groups of 16 as in Exp. I. Two *S*s were discarded because of apparatus failure and 5 were discarded because on three consecutive trials they failed to trip the first treadle within the 2-min. time limit or failed to trip one of the second treadles within the 3-min. limit. (On such trials, *S* was placed in a randomly selected goal box on the first occasion and subsequent odd-numbered occasions, and in the alternative goal box on all even-numbered occasions. The time score was recorded as 180 sec. and the goal box into which *S* was placed was recorded as chosen.) These discarded *S*s were replaced by *S*s of the same description. The apparatus was the same as that used in Exp. I.

Prehandling.—Prehandling was the same as in Exp. I, but the feeding schedule was changed slightly. The once-daily ration of 10 gm. was presented 2 hr. after the start of prehandling or experimental training, since the altered procedure increased the total time of experimental training each day from 15–25 min. to 90–110 min.

Experimental training.—Except for a change in intertrial interval and magnitude of reward, all details were the same as in Exp. I. In Exp. II, 3–4 min. elapsed between the placement of *S* in its carrying cage following a trial and its removal at the beginning of the subsequent trial. In addition, only one 45-mg. pellet was present in the proper goal box on each trial. These changes applied to acquisition, overtraining, and reversal.

Acquisition.—Fifteen free trials were given on each of 2 consecutive days. Groups Fr, Co, and In all received acquisition on Days 1 and 2 of experimental training. Group N was prehandled (in the same manner as before) on Days 1–9 and then given acquisition on Days 10–11.

Overtraining.—Group Fr was given 20 free trials on each of Days 3–6 and 8–9 and 15 free trials on each of Days 10–11, for a total of 150 free overtraining trials. Group Co was given 20 trials, all forced to the correct side, on each of Days 3–6 and 8–9 and 15 trials, also forced correct, on Days 10–11, for a total of 150 forced correct trials. Group In was given 40 trials on each of Days 3–6 and 8–9 and 30 trials on Days 10–11. Half of these 300 trials were forced correct and half forced incorrect.

Reversal.—On Days 12 and 13 all *S*s were given reversal training as in Exp. I.

Results

Acquisition.—The mean proportion of correct choices on the last five trials of acquisition was .93. An analysis of variance of the last half of acquisition revealed no significant difference among the groups ($F < 1$). To check comparability with one previous study using spatial discrimination, it was determined for each *S* on what trial, if at all, it reached the criterion defined in Exp. I. For the four groups combined, 35 of the 64 *S*s reached this criterion, the median for completion of the criterion being Trial 29. This means that by Pubols' definition the median *S* received one trial of overtraining during acquisition.

Overtraining.—Group Fr made a a total of 2.7% incorrect choices during overtraining. The median number of correct responses after the last error was 100.

Reversal.—The course of reversal learning for the four groups is shown in Fig. 1, and the mean number of

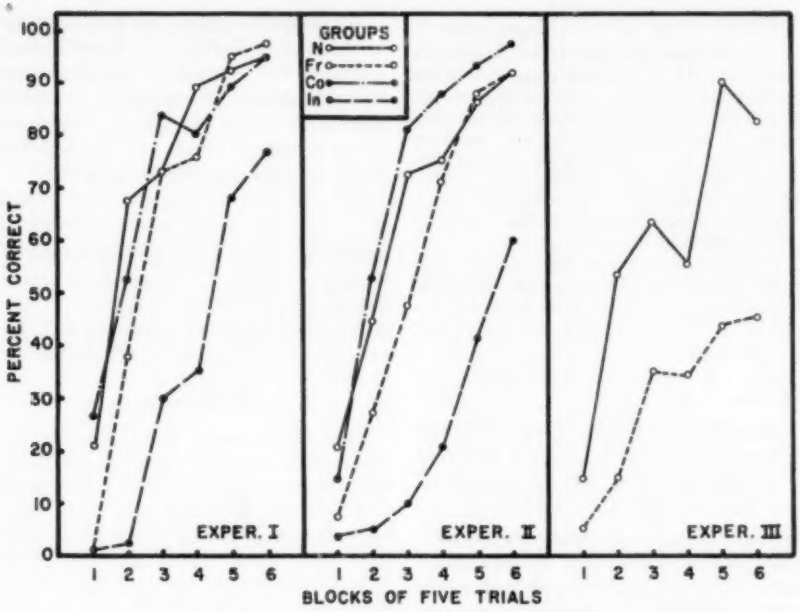

FIG. 1. Reversal learning curves in the three experiments.

correct choices for all 30 trials is shown in Table 1. The pattern of group differences is clearly similar to that found in Exp. I. Analysis of variance of the four group means yielded an F of 22.77, significant at the .001 level. By t test, Group In was found to be lower than each of the others at the .001 level, while Group Fr was lower than Group Co at the .05 level. Other differences were not significant.

The mean number of errors before the first reversal is shown in Table 2. Group N does not differ significantly from Groups Fr and Co, but all other differences are significant at least at the .05 level.

The reversal data were also analyzed according to the same criterion as in Exp. I. The numbers of *Ss* reaching criterion in the four groups were 8, 7, 14, and 1, respectively, yielding a χ^2 of 21.3, significant at the .001 level for 3 *df*.

Speeds.—Speeds on the first five reversal trials are shown in Table 3. The overall difference among the groups is significant at the .01 level both for stem speeds ($F = 7.27$) and for total speeds ($F = 5.95$). In both cases the two higher groups differ significantly from the two lower, but Group Fr does not differ significantly from Co or Group N from In.

EXPERIMENT III

Since the pattern of results in Exp. II did not differ appreciably from that in Exp. I, a further attempt was made to find an ORE. Since *Ss* in Exp. I and II were younger than those in some of the previous studies that found an ORE, older *Ss* were used in Exp. III. Only Groups N and Fr were included in this experiment.

Method

Subjects and apparatus.—The *Ss* were 27 experimentally naive female albino rats of the Sprague-Dawley strain, 120–121 days of age at the beginning of experimental training. Three *Ss* were discarded because on three consecutive trials they failed to trip the treadles within the specified time limits, which were the same as in Exp. II. These discarded *Ss* were replaced by other *Ss* of the same description.

Procedure and design.—The *Ss* were randomly assigned to one of two groups which were treated exactly like Groups N and Fr of Exp. II. There were 13 *Ss* in Group N and 14 *Ss* in Group Fr.

Results

Acquisition.—Mean proportion of correct choices on the last five trials of acquisition was .94 for Group N and .84 for Group Fr. A t test on the number of correct responses in the latter half of acquisition revealed that Group N was significantly superior at the .02 level ($t = 2.63$). No reason for this difference is apparent. The median S in the two groups combined just reached criterion, with no additional trials.

Overtraining.—Group Fr had a mean of 2.0% incorrect choices during overtraining. The median number of correct responses after the last error was 102.

Reversal.—The course of reversal learning is shown in Fig. 1. The total numbers of correct responses in reversal are given in Table 1. The superiority of Group N is significant at the .01 level ($t = 2.89$). In view of the superiority of Group N in acquisition, this might be attributed to a chance superiority in learning ability of *Ss* in the group. However, the within-groups correlation between number of correct choices in the latter half of acquisition and number correct during reversal is negative and nonsignificant ($r = -.22$), which makes such an interpretation implausible.

Mean errors before the first reversal are shown in Table 2. The difference between the groups is significant at the .01 level. The reversal criterion previously described was reached by 6 *Ss* in Group N and 3 in Group Fr, yielding a nonsignificant χ^2 of .91.

Speeds.—Stem speeds and total speeds for the two groups are shown in Table 3. As in both previous experiments, Group Fr was faster than Group N on both measures. The differences were not significant, however (t's = .85 and 1.20, respectively).

Discussion

Reversal learning was found to be temporarily retarded by free-trial overtraining, greatly retarded by forced-trial overtraining when half the trials were to each side, and unaffected by forced-trial overtraining when all the trials were to the correct side. These effects were primarily accounted for by the duration of perseveration on the formerly correct side in the different groups at the beginning of reversal, as may be seen by comparing Tables 2 and 3.

The most noteworthy aspect of these findings is the failure to confirm the finding of other investigators that overlearning facilitates reversal. We seem to have ruled out the possibility that the discrepancy between our results and those of previous *Es* is due to age of *Ss*, number of overlearning trials, or intertrial interval. Four differences remain between our procedure and earlier ones.

1. We used a constant number of acquisition trials before overtraining instead of carrying all *Ss* to a criterion. Because of the variability of the *Ss*, a training procedure which gives a constant number of trials may fail to insure that all *Ss* will master the original task or may permit a few to receive a degree of overtraining. Either of these conditions might facilitate reversal performance of Group 1 and reduce the chance of demonstrating an ORE. For this reason comparisons were made among three subgroups of Groups N and Fr:

(*a*) those that failed to reach the acquisition criterion, (*b*) those that had 5 or fewer acquisition trials after reaching criterion, and (*c*) those that had 6 to 10 (the maximum possible) acquisition trials beyond criterion. In all three subgroups in all three experiments, Group N made more correct reversal responses than Group Fr. (In view of the small *N* in these subgroups, no statistical tests were made.) This suggests that the failure to use an acquisition criterion was not critical. There might be an interaction, however, between this consideration and the next one.

2. We gave a constant number of reversal trials rather than carrying each *S* to a reversal criterion. Perhaps *S*s in Group Fr, once they overcame their initial disadvantage, would have surpassed Group N in reaching criterion. Neither the reversal learning curves nor the number of *S*s reaching criterion in the two groups lend any support to this view. As for the possible interaction of Differences 1 and 2, of the *S*s in all three experiments that reached the acquisition criterion with five or fewer trials to spare, 5 in Group N and 4 in Group Fr reached the reversal criterion.

3. We tried to control differences in amount of handling between overtrained and nonovertrained groups by giving Group N extra handling when they were not being run. This may have controlled away the whole effect. If this should be the explanation, it would make the overlearning-reversal effect a more trivial phenomenon than has been suspected.

4. Our task was a simple spatial discrimination with no irrelevant visual cues (such as Pubols had) and no separation of place and response cues (as with Brookshire et al.). If observing responses are the crucial factor in the overlearning-reversal effect, such a simple discrimination may be too easy for such responses to be important. This explanation implies that our task is easier than those in which the effect has been found. In terms of acquisition rate, this appears to be true of most of the other studies, though not of Pubols' Exp. 2. If this exception can be explained by the extensive pretraining in Pubols' experiment leading to faster learning, then the simplicity of our discrimination might well be the crucial factor in our failure to find an overlearning-reversal effect. This would be consistent with D'Amato and Jagoda's (1962) failure to find an overlearning-reversal effect in a spatial discrimination. Their study, of which we were unaware when conducting ours, is to our knowledge the first published failure to find an overlearning-reversal effect with high overlearning. The situation is complicated, however, by an unpublished study of Erlebacher (1961). He failed to find the overlearning-reversal effect, even though he used a brightness discrimination. The necessary conditions for obtaining an overlearning-reversal effect thus remain in doubt.

Our results and the strikingly similar findings of D'Amato and Jagoda seem to be clearly incompatible with the first three interpretations of discrimination reversal listed in the introduction. These findings can, however, be reconciled with Interpretations 4 and 5, both of which are concerned with avoidance of the formerly incorrect side. Either of these interpretations is consistent with the rank order of our three overtraining groups. The unexpected superiority of Group N in reversal could then be explained by their weaker tendency to approach the formerly correct side, a tendency supported by the generally slower speeds of Group N. The similar patterns of stem speed and total speed suggest that time spent in the stem rather than time spent in the choice area or arms was the major source of variance in both speed measures.

Summary

Experiment I attempted to find the cause of the overlearning-reversal effect by comparing T maze reversal learning by four groups of rats that received different patterns of overtraining in acquisition. Reversal was fastest for the group receiving no overtraining and the group receiving all its overtraining trials forced to the correct side. Free-choice

overtraining gave somewhat slower reversal, and overtraining with an equal number of forced trials to the two sides gave much slower reversal.

In view of the failure to replicate previous findings of faster reversal after overtraining, two further experiments were run in an attempt to replicate these earlier findings. Both experiments gave the same pattern of results as Exp. I; no overlearning-reversal effect was found. These results appear to be consistent with interpretations in terms of stimulus satiation or of avoidance of nonrewarded cues, but not with interpretations in terms of observing responses, discriminability, or frustration.

References

Birch, D., Ison, J. R., & Sperling, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

Brookshire, K. H., Warren, J. M., & Ball, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54**, 98–102.

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

Cotton, J. W., Lewis, D. J., & Metzger, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

D'Amato, M. R., & Jagoda, H. Analysis of the role of overlearning in discrimination reversal. *J. exp. Psychol.*, 1961, **61**, 45–50.

D'Amato, M. R., & Jagoda, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, **64**, 117–122.

Erlebacher, A. Reversal learning in rats as a function of percentage reinforcement and degree of overlearning. Unpublished doctoral dissertation, University of Wisconsin, 1961.

Glanzer, M. Curiosity, exploratory drive, and stimulus satiation. *Psychol. Bull.*, 1958, **55**, 302–315.

North, A. J., & Clayton, K. N. Irrelevant stimuli and degree of learning in discrimination learning and reversal. *Psychol. Rep.*, 1959, **5**, 405–408.

North, A. J., & Stimmel, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

Pubols, B. H., Jr. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

(Received November 14, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 541-544

MONETARY INCENTIVE AND RANGE OF PAYOFFS AS DETERMINERS OF RISK TAKING[1]

LEONARD KATZ

University of Massachusetts

Myers and associates (Myers & Fort, 1961; Myers & Katz, 1962; Myers & Sadler, 1960) have been concerned with the effects of parameters of the payoff distribution upon the choice between gambling and not gambling. Myers and Sadler (1960) varied the number of chips which might be won or lost on each gamble (range), the average payoff being zero. When the alternative to gambling was the sure gain of one chip, gambling increased with increases in range; when the alternative to gambling was the sure loss of one chip, gambling decreased with an increase in range.

In the present experiment chips worth no money and chips worth 5¢ were used to provide data on effects of incentive value. Monetary incentive and range of payoffs are similar to each other in that an increase in either increases the risk associated with each gamble. The implied hypothesis is that increased incentive, like increased range, may result in more gambling, where the alternative to gambling is the sure gain of one chip, and in less gambling when the alternative to gambling is the sure loss of one chip. The objective of the present study was, therefore, to obtain data to test this hypothesis. The *S*'s decision to gamble under different payoff ranges was followed by the loss or gain of chips which had no monetary value or was followed by the loss or gain of chips worth 5¢.

In addition to providing data on the effects of monetary incentive, the present study investigated the effects of a payoff range greater than that used in the Myers and Sadler (1960) study. To control for any range effects due to differences among ranges in sequences of payoffs, one sequence was randomly generated and the other two were derived from it in a manner described in the procedure section.

METHOD

Materials.—Four decks of 100 3 × 4 in. white cards were prepared. The known payoff deck contained 50 cards with +1 written on them, alternated randomly with 50 cards with −1. The other decks, of 100 cards each, provided for three different ranges of unknown payoffs. The narrow-range deck (N), had integers randomly chosen from +2 to +6 and from −2 to −6. The medium range deck (M), was constructed by adding 10 to every positive number of (N) and subtracting 10 from every negative number, giving a range of +12 to +16 and −12 to −16, retaining the same ordinal positioning of cards as found in Deck N. The wide range deck (W), was constructed by adding and subtracting 20 to the integers of Deck N.

Procedure.—On each of 3 successive days *S* was presented with the known payoff deck, and a different one of the three unknown payoff decks. Order of use of Decks N, M, and W was counterbalanced in a 3 × 3 Latin square with 6 *S*s given each order. Eighteen *S*s gambled only for poker chips (0¢) and 18 others gambled for chips worth 5¢ apiece. The main features of the experimental design are shown on the left-hand side of Table 1.

The *S*s were tested individually, being given full instructions at the beginning of Session 1. For Sessions 2 and 3 *S*s were told

[1] This research was supported by National Science Foundation Grant G-11380 and National Institute of Mental Health Grant M-3803, and was part of a study submitted in partial fulfillment of the Master of Science degree at the University of Massachusetts. The author is indebted to Jerome L. Myers, who served as thesis adviser.

TABLE 1

MEANS OF PROPORTIONS OF GAMBLING RESPONSES ON +1 AND −1 TRIALS FOR SUCCESSIVE 25-TRIAL BLOCKS

Incentive	Alternative to Gambling	Range	Trials				Overall Trials
			1–25	26–50	51–75	76–100	
0¢	+1	N	.594	.539	.456	.511	.525
		M	.620	.551	.494	.604	.567
		W	.466	.472	.394	.596	.482
	−1	N	.903	.910	.874	.911	.899
		M	.796	.796	.822	.744	.798
		W	.838	.836	.756	.800	.797
5¢	+1	N	.654	.588	.489	.563	.573
		M	.688	.537	.461	.574	.565
		W	.658	.574	.483	.689	.601
	−1	N	.833	.893	.881	.878	.871
		M	.796	.790	.815	.761	.790
		W	.782	.756	.741	.750	.757

only that the unknown payoff deck of that day was a new one. Two hundred poker chips were stacked in front of *S*, and *S*s with a monetary incentive (5¢) were told that each chip was worth 5¢; i.e., their initial stake was worth $10.00.

Briefly, *S*s were instructed to turn over the top card in the deck of known payoffs at the beginning of each trial. They then chose between standing pat by accepting the gain or loss of one chip represented by the card and gambling by drawing the top card from the deck of unknown payoffs. If *S* decided not to gamble, the top card in the unknown payoff deck was turned anyway, showing what he would have won or lost had he gambled.

Ratings.—At the end of each session, *S*s were given an 11-point scale from −5 to +5, along which they rated the means of the known payoff deck and of the unknown payoff deck used that session. They were also asked to describe their gambling strategies and changes in strategy. After Session 3, *S*s were asked what they would have done differently in Sessions 1 and 2.

Subjects.—The *S*s were 36 male undergraduates enrolled in the university summer session, who were divided randomly into two groups of 18 *S*s each. Each *S* was paid $3.00.

RESULTS

Choices.—The scores were proportions of choices to gamble on both +1 and −1 trials during each of four 25-trial blocks. For example, if in a particular block *S* gambled on 3 of the 12 (or 13) trials in which the alternative to gambling was +1, his +1 proportion for that block was .250 (or .231). Table 1 presents means of these proportions for each combination and range. An analysis of variance was performed on arc-sine transforms of the proportions. The results of this analysis (shown in Table 2), in conjunction with the relationships shown in Table 1, suggest the following conclusions: (*a*) under all combinations of range and incentive, more risk-taking occurs on −1 trials than on +1 trials ($P < .001$); (*b*) the difference in incentives had little effect on total number of risks taken; (*c*) more gambling occurred with Deck M than with Deck N on +1 trials and the reverse on −1 trials ($P < .01$); and (*d*) value, incentive, and range have a joint effect on gambling ($P < .001$). A further breakdown suggests that this effect is due largely to differences in the quadratic curvature of the V-I

curves over the three ranges. For +1 trials, the curves for incentive are essentially mirror images of each other; for −1 trials, the curves for incentive are essentially parallel to each other.

In addition, there were a number of significant interactions involving blocks of trials. These effects are probably due to the interaction of range, incentive, and value with the effects of both the temporal sequence of blocks within a single session and the different mean payoffs of blocks.

Ratings.—Table 3 shows the means of *Ss*' ratings of the means of the known and unknown payoff decks for each combination and incentive. The means for the known payoff deck were closer to the true mean of zero and were less variable than the means for the unknown payoff decks. Neither set of means varied systematically with incentive or range of unknown payoff. Nor were there any systematic relationships between these deck means and either overall proportions of risks or proportions of risks for the last 25 trials. Most *Ss* reported following a "gambler's fallacy" [2] strategy and none reported an awareness of gambling differentially with different decks of unknown payoffs. Incentive had no differential effect.

TABLE 2

ANALYSIS OF VARIANCE OF ARC-SINE TRANSFORMS OF THE PROPORTION OF GAMBLING RESPONSES IN EACH TRIAL

Source	*df*	*MS*	*F*
Incentive (I)	1	710.00	
Ss/I	34	2,047.71	
Value (V) (known payoff)	1	103,825.00	33.72***
Range (R)	2	1,361.40	2.22
Blocks (B)	3	1,454.93	4.90**
V × B	3	987.23	3.28*
V × R	2	1,855.60	8.73**
R × B	6	187.01	8.87**
R × B × V	6	367.75	2.25*
V × I	1	2,910.50	1.07
B × I	3	56.80	0.19
R × I	2	292.70	0.47
V × I × R	2	1,606.10	13.66***
R × I × B	6	61.66	2.82*
V × I × B	3	238.16	1.39
R × B × V × I	6	154.27	1.57
Ss × V/I	34	2,701.36	
Ss × B/I	102	296.53	
Ss × R/I	86	611.05	
Ss × V × B/I	102	300.18	
Ss × V × R/I	68	210.20	
Ss × R × B/I	204	21.80	
Ss × V × B × R/I	204	162.91	

* $P < .05$.
** $P < .01$.
*** $P < .001$.

TABLE 3

MEANS OF *Ss*' RATINGS OF DECK MEANS

Range	Payoff			
	Known		Unknown	
	0¢	5¢	0¢	5¢
N	.33	.16	.44	−0.66
M	.33	.16	.00	1.05
W	.44	−.22	.88	0.88

DISCUSSION

The decision to gamble or not to gamble following known outcomes of loss or gain of a poker chip was investigated as a function of three ranges of unknown payoff involving the loss or gain of chips worth nothing or worth 5¢. Previous findings (Myers & Katz, 1962; Myers & Sadler, 1960) of more gambling when the alternative to gambling was the loss of a chip than when the alternative was the gain of a chip were confirmed. This difference was reduced by monetary incentive, although the interaction was not statistically significant: trials when the alternative was −1 were followed by fewer choices to gamble for chips worth 5¢ than for those worth nothing, while trials where the alternative was +1 were followed by more choices to gamble for chips worth 5¢ than for those

[2] Due to the procedure for constructing the payoff sequences, prediction of the alternative event following a run is not really fallacious.

worth nothing. Thus, there is some suggestion that monetary payoff and range are functionally equivalent; for all three ranges monetary incentive yielded an increase in gambling on +1 trials, a decrease on −1 trials.

Other experiments have shown the form of value and range interaction obtained for the 5¢ incentive group. Myers and Sadler, using three ranges up to and including the range of Deck M of the present study, found that gambling consistently increased as range increased when the alternative was a gain of one chip, but gambling decreased as range increased when the alternative was a one-chip loss. Myers and Katz (1962) obtained similar results through a range of Deck M. Suydam and Myers (1962), using a very different procedure, found this convergence of positive and negative value curves over range, for several values. It may be assumed that the gamble represents an approach-avoidance conflict, both tendencies increasing as range does. Within this frame of reference, the data suggest that, as range increases, the avoidance gradient rises more swiftly against negative alternatives; the approach gradient rises more swiftly against positive alternatives.

Summary

The effects on gambling behavior of monetary incentive, range of payoffs for gambling, and value of a payoff which could be taken in lieu of gambling were determined. One group of 18 *S*s gambled for chips only and another group of 18 *S*s gambled for chips worth 5¢ each. Three ranges of unknown payoffs were used, one at each of three sessions. The known payoff, the acceptance of which was the alternative to gambling, remained constant.

Neither incentive nor range had a significant effect upon the total number of risks taken. The *S*s gambled significantly more when the alternative to gambling was a loss (−1) than when it was a gain (+1). Several interactions were significant, including Range × Value and Range × Value × Incentive. These led to the conclusion that gambling is affected differentially on +1 and −1 trials by the range of chips to be gained or lost, and by the interaction of range and monetary incentive.

References

Myers, J. L., & Fort, G. G. A sequential analysis of gambling behavior. Paper presented at Psychonomic Society, Columbia University, 1961.

Myers, J. L., & Katz, L. Range of payoffs and feedback in risk-taking. *Psychol. Rep.*, 1962, **10**, 483–486.

Myers, J. L., & Sadler, E. Effects of range of payoffs as a variable in risk taking. *J. exp. Psychol.*, 1960, **60**, 306–309.

Suydam, M. M., & Myers, J. L. Two parameters of risk-taking behavior. *Psychol. Rep.*, 1962, **10**, 559–562.

(Received November 14, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 545–548

SPATIAL S-R CONTIGUITY IN HUMAN DISCRIMINATION LEARNING

C. D. STANDISH AND R. A. CHAMPION

University of Sydney

Though differing in purpose, both of the experiments to be reported are basically similar in design to an earlier study (Champion & Standish, 1960) in which preliminary training was given with two pairs of stimuli in fixed spatial relations (spatial discrimination learning) followed by a test stage involving the two pairs and their transposes (nonspatial learning). Under these conditions negative transfer from training to test occurred when between-pair differences among the stimuli were greater than within-pair differences, but positive transfer occurred when within-pair differences dominated. These findings were interpreted in terms of the type of S-R theory proposed by Spence (1960); in particular, the differential transfer was explained through the presence or absence of within-pair discriminations on the part of *S* in the training period, it being argued that the occurrence of these discriminations promoted positive transfer while their absence caused interference. The basis of the present studies rested in the further assumption that spatial S-R contiguity would also promote within-pair discriminations, and it was predicted that groups trained and tested under such conditions would show positive transfer, even though between-pair differences were dominant.

The first experiment consisted of a repetition of the 1960 study with the exception that *S* was required to respond by pressing a button adjacent to the stimulus rather than one some distance from it. In order to make a severe test of the effect of this form of S-R contiguity, the stimuli in each pair were presented close together in space (S-S proximity), but this allowed the possibility that the result of the experiment was due to proximity rather than contiguity, the two factors being confounded. The second experiment was therefore conducted in an attempt to separate these two factors and attention was concentrated on the contiguity variable, with degree of proximity held constant.

EXPERIMENT I

Method

Subjects.—The *S*s were 44 undergraduates from courses in psychology at the University of Sydney, there being 13 *S*s in Groups 1 and 2, and 9 in Groups 3 and 4. Color-blind students were excluded.

Apparatus and stimuli.—The apparatus was identical with that used previously (Champion & Standish, 1960), allowing *S* to be presented with colored circles of light 1 in. in diameter, except that the lights were only ¼ in. apart in the horizontal midline of the 5 × 2 in. milk-glass screen, and the response buttons were mounted as close to the stimuli as possible, one button being located immediately above each light. In an attempt to equate within-pair differences the stimulus settings were adjusted to the following values of red (R) and green (G) on the Munsell system, with brightness and saturation held approximately constant: Stimulus R_1 = 5R, R_2 = 5YR, G_1 = 7.5GY, G_2 = 2.5GY. The results of the 1960 study had already been confirmed with these settings (Standish, 1960).

Procedure.—The instructions to *S* were slightly modified to allow for the new location of the response buttons, and between responses *S*'s hand rested on the table immediately below the vertical midline of the screen, but otherwise the procedure was identical with that reported earlier, the control groups being denied preliminary training

TABLE 1

STIMULUS PAIRS FOR *Ss* IN EXP. I

Stage	Experimental Groups		Control Groups	
	1 b > w	3 w > b	2 b > w	4 w > b
1	$R_1+:R_2-$ $G_2-:G_1+$	$G_1+:R_2-$ $G_2-:R_1+$	—	—
2	$R_1+:R_2-$ $G_2-:G_1+$ $R_2-:R_1+$ $G_1+:G_2-$	$G_1+:R_2-$ $G_2-:R_1+$ $R_2-:G_1+$ $R_1+:G_2-$	$R_1+:R_2-$ $G_2-:G_1+$ $R_2-:R_1+$ $G_1+:G_2-$	$G_1+:R_2-$ $G_2-:R_1+$ $R_2-:G_1+$ $R_1+:G_2-$

Note.—The spatial relations of the symbols in each pair correspond to the actual left-right arrangements of the stimuli. The designations b > w etc. refer to relative magnitudes of stimulus differences between and within pairs.

so as to allow an assessment of the nature and amount of transfer. A summary of the procedure and stimulus pairs is given in Table 1.

Results

The results, in terms of mean trials to the criterion of eight successive correct trials, are summarized in Table 2. Because of the presence of some extreme scores, nonparametric statistical tests were used. The application of a *U* test to the data of the first stage for Groups 1 and 3 (two stimulus pairs in fixed spatial relations) confirmed the earlier finding of superior performance with between-pair differences dominant ($U = 6$ with $n_1 = 13$ and $n_2 = 9$, $P < .001$).

TABLE 2

TRIALS TO CRITERION IN EXP. I

Stage	Between-Pair Differences Dominant		Within-Pair Differences Dominant	
	Group 1 (Exp.)	Group 2 (Control)	Group 3 (Exp.)	Group 4 (Control)
Training				
Mean	14.4	—	25.9	—
SD	3.6	—	9.4	—
Test				
Mean	19.0	32.8	16.3	33.6
SD	22.8	29.4	4.8	19.8

The use of a factorial median test (Sutcliffe, 1957) with the data of the test stage (two stimulus pairs and their transposes) failed to reveal any differential transfer, there being no significant interaction between the experimental-control variable and type of stimulus difference ($\chi^2 = 3.38$ for 1 *df*). Subsequent *U* tests showed that significant positive transfer occurred with both between-pair and within-pair differences dominant ($U = 38.5$ with n_1 and $n_2 = 13$, $P < .02$, and $U = 11.5$ with n_1 and $n_2 = 9$, $P < .02$, respectively).

EXPERIMENT II

Method

Subjects.—The *Ss* were 60 undergraduates from courses in psychology, divided into four groups of 15. Color-blind students were not excluded.

Apparatus.—The apparatus consisted of a sheet of building board 72 in. wide and 36 in. high, mounted vertically on a table in front of *S* and containing two $2\frac{1}{4} \times 2\frac{1}{4}$ in. pearl-perspex squares 4 in. apart in the horizontal midline. An attempt was made to locate *S* equidistant from the two squares so that the centers subtended the same visual angle as did the centers of the circles in the 1960 study. Housed behind the perspex squares were 15-w. lamps whose intensity was controlled by two variacs. The duration of illumination of the stimuli on any trial was 200 msec. In order to achieve spatial contiguity of stimulus and response *S* was instructed to use response buttons mounted on each perspex square, one in each top corner nearer to the vertical midline of the building board. To obtain noncontiguity conditions, *S* was told to use two buttons mounted side by side in a small metal box located 3 in. in front of the board, the box being centered 16 in. to the right of the midline of the board. The response buttons activated indicator lamps as a signal to *E*.

Stimuli.—The chief basis of discrimination in Exp. II was intensity of illumination, and the four values of the stimuli in foot-Lambert units were as follows: $B_1 = 125$, $B_2 = 100$, $D_1 = 2.5$, $D_2 = 3.2$. These illumination intensities were measured with a photometer at the usual location of *S*'s head. Lights B_1 and B_2 were similar bright stimuli, whereas D_1

and D_2 were similar dull stimuli. The change from bright to dull illumination of the filament lamps was accompanied by the usual change in hue.

Procedure.—The instructions to *S* were again modified to the extent demanded by the changes in the apparatus and the conditions of training. The independent variable was degree of spatial S-R contiguity (buttons near to or far from lights) and the stimulus conditions were limited to dominance of between-pair differences, the required discriminations being between B_1 and B_2, and between D_1 and D_2. Otherwise the procedure was as before, with training and test stages for the two experimental groups (Group 1, contiguity, and Group 3, noncontiguity), and test stages alone for the corresponding control groups (Groups 2 and 4, respectively). The criterion of learning was increased to 12 successive correct trials.

Results

The trials-to-criterion scores for the training and test stages are presented in Table 3. For these data it was possible to use a parametric statistic, and the application of a *t* test to the training scores of Groups 1 and 3 revealed no significant difference. An analysis of the variance for the test data of the second stage showed that while the pretraining had no consistent effect (experimental vs. control groups), the influence of spatial S-R contiguity and degree of differential transfer (interaction) were both statistically significant. Follow-up *t* tests showed that the differences between Groups 1 and 2, and Groups 1 and 3 were significant ($t = 2.56$, $P < .02$, and $t = 4.74$, $P < .01$, respectively, for 28 *df*) but that the differences between Groups 2 and 4, and Groups 3 and 4 were not significant. It thus appears that positive transfer occurred only under contiguity conditions, but that there was no negative transfer with noncontiguity of S and R.

TABLE 3

TRIALS TO CRITERION IN EXP. II

Stage	Contiguity		Noncontiguity	
	Group 1 (Exp.)	Group 2 (Control)	Group 3 (Exp.)	Group 4 (Control)
Training				
Mean	25.6	—	29.7	—
SD	20.6	—	20.4	—
Test				
Mean	18.9	45.7	68.5	57.6
SD	8.9	29.4	31.3	29.9

DISCUSSION

The results of the two experiments lead to the conclusion that spatial S-R contiguity, like between-pair similarity, leads to positive transfer from the spatial to the nonspatial learning situation (training to test), and if the theoretical interpretation given in terms of S-R theory is correct, then this is because the contiguity promotes more effective within-pair discrimination in the first stage. The chief finding of Exp. I was that positive transfer from training to test may be obtained even when between-pair differences are dominant, provided that *S* be required literally to approach the positive stimulus cue in the course of the response. This result is to be contrasted with that of the 1960 study, where the dominance of between-pair differences produced negative transfer. That the present outcome cannot be attributed to the greater proximity of the stimulus cues in Exp. I is demonstrated by the similar result in Exp. II with a return to the same degree of nonproximity as obtained in the 1960 study.

The data of Exp. II suggest that the direct effects of contiguity may not be as powerful with adult humans as with children and lower animals, by comparison with the marked effects obtained by Murphy and Miller (1958, 1959). Learning was more efficient under contiguity conditions both in the training of Groups 1 and 3 and in the test periods of Groups 2 and 4 (Table 3), but in neither case was the difference statistically significant. However, when the two scores for each *S* in the experimental

groups of Exp. II (Groups 1 and 3) on the training and test stages were combined, so as to make the two tasks one, a significant difference emerged ($t = 4.32$ for 28 *df*, $P < .01$), possibly due to the greater stability of the scores. It will also be noted that the significant negative transfer obtained in the 1960 study with noncontiguity and between-pair differences dominant was not duplicated in Groups 3 and 4 of Exp. II, although the result was in the same direction.

Spence's recent elaboration of an S-R theory of selective learning (Spence, 1960) points to some of the complexities to be coped with in any detailed formulation, but brief consideration may profitably be given to the broader theoretical significance of spatial S-R contiguity. In the present context this variable has been assumed to act through within-pair discrimination, and in S-R terms the efficiency of this latter discrimination depends in turn upon the relative strengths of the orienting and approach tendencies to the positive and negative discriminanda. The necessary link between spatial contiguity and the relative strengths of the correct and incorrect responses may now be provided if it be allowed that spatial contiguity amounts to temporal contiguity of stimulus and response, for the importance of the latter variable in simple learning seems undoubted (e.g., Champion, 1962). The equation of spatial and temporal contiguity in the discrimination-learning situation follows from the fact that when *S* is required to respond by pressing a button in or near the positive stimulus (contiguity) then there is more likely to be a short time interval between exposure to the stimulus and the occurrence of the response than if *S* has to look away from the stimulus to locate the appropriate response button (noncontiguity). This effect might show up more clearly if the stimuli were presented for a long time interval, but the argument applies even with intervals as short as 100 msec., for there is no requirement that stimulus and response overlap in time, the terms "contiguity" and "noncontiguity" being relative rather than absolute.

Summary

Two experiments on discrimination learning were conducted under conditions in which preliminary training was given with two pairs of stimuli in fixed spatial relations, followed by test learning involving the two pairs and their transposes. The aim was to test the effects of spatial S-R contiguity, for it was predicted that contiguity would cause positive transfer from training to test. The prediction was confirmed, and the result was interpreted in S-R terms with the hypothesis that contiguity, like between-pair similarity, promotes within-pair discriminations on the part of the learner, it being supposed that spatial contiguity has this effect through the more basic variable of temporal contiguity.

References

Champion, R. A. Stimulus-response contiguity in classical aversive conditioning. *J. exp. Psychol.*, 1962, **64**, 35–39.

Champion, R. A., & Standish, C. D. Stimulus differences in discrimination learning. *J. exp. Psychol.*, 1960, **60**, 78–82.

Murphy, J. V., & Miller, R. E. The effect of the spatial relationship between the cue, reward, and response in simple discrimination learning. *J. exp. Psychol.*, 1958, **56**, 26–31.

Murphy, J. V., & Miller, R. E. Spatial contiguity of cue, reward, and response in discrimination learning by children. *J. exp. Psychol.*, 1959, **58**, 485–489.

Spence, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Standish, C. D. Stimulus differences and response tendencies in discrimination learning. Unpublished master's thesis, University of Sydney, 1960.

Sutcliffe, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychol. Bull.*, 1957, **54**, 134–137.

(Received December 4, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 549–550

SUPPLEMENTARY REPORT: STIMULUS FAMILIARIZATION IN PAIRED-ASSOCIATE LEARNING

RUDOLPH W. SCHULZ AND IRVING F. TUCKER

State University of Iowa

When *Ss* are familiarized with the stimulus units but not the response units of a list prior to paired-associate (PA) learning, it has generally been found that stimulus familiarization (SF) has a slight inhibitory effect or no effect at all on PA performance (Underwood & Schulz, 1960). Recently, Gannon and Noble (1961) have reported significant facilitation of performance on a list of paired dissyllables following 20 trials of SF. Since the latter result is the one to be expected if frequency of prior experience is the vehicle through which stimulus meaningfulness has its effect on PA performance, it is a result with considerable theoretical significance (e.g., Cieutat, Stockwell, & Noble, 1958; Underwood & Schulz, 1960). However, the finding of a positive effect from SF is also conspicuously inconsistent with the results of previous studies. Therefore it seemed especially important to further assess the reliability of this result and consider potential alternative explanations for it.

One such alternative is that Gannon and Noble's procedure of having *S* articulate the stimulus unit during the PA anticipation interval, when combined with variation in amount of SF, may have inadvertently produced simultaneous variation in the *effective length* of the anticipation interval. Since practice in articulation of stimulus units is directly related to amount of SF, familiarized *Ss* might spend relatively less of the 2-sec. anticipation interval for stimulus articulation than nonfamiliarized *Ss*. On the basis of the presumed direct relationship between PA performance and length of the anticipation interval Gannon and Noble's results would then be expected. The present experiment tested this hypothesis by comparing the performance of *Ss* instructed to pronounce the stimulus units during PA anticipation with performance of *Ss* instructed *not* to pronounce. A significant interaction of PA instruction and amount of SF will be required to support the present contention.

Method.—A 2 × 3 factorial design with 2 levels of PA instruction—articulation (A) vs. nonarticulation (NA) of stimulus units—and 3 amounts of SF (0, 20, and 60 trials) was used. The six respective conditions will be referred to in terms of the values of the independent variable associated with them (e.g., Cond. A_0 articulation instructions and O familiarization, Cond. NA_{60} nonarticulation instructions and 60 trials of familiarization, etc.).

The materials and procedures were identical to those used by Gannon and Noble (1961) with the following exceptions: (*a*) A .85-sec. rate of presentation was used during familiarization; (*b*) PA performance consisted of 17 anticipation trials.

A total of 144 *Ss*, 24 per condition, taking introductory psychology at the University of Iowa were randomly assigned to conditions as they appeared at the laboratory. The *Ss* had not served in prior verbal learning experiments.

Results and discussion.—Performance on the PA list under the six conditions, in terms of mean total number of correct responses during 17 anticipation trials, is shown in Fig. 1. The predicted interaction between PA instructions and amount of SF was obtained, and is shown by an analysis of variance to be the only significant effect ($F = 3.78$, $df = 2/138$, $P < .05$). In agreement with the results of Gannon and Noble (1961), performance was a monotonic increasing function of number of familiarization trials when *Ss* were required to pronounce the stimulus terms of a PA list prior to anticipation of the response terms. However, with

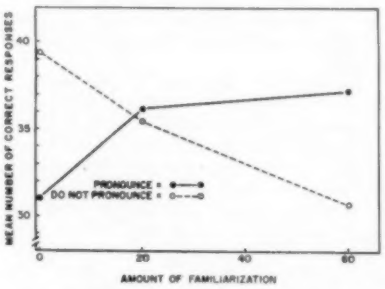

FIG. 1. Mean total number of correct responses during 17 anticipation trials as a function of PA instructions and number of stimulus familiarization trials. (The standard error of the means in Fig. 1, as estimated from the within-groups *MS* of the overall analysis of variance, was 2.71.)

nonarticulation PA instructions, performance was inversely related to amount of familiarization. It can also be seen from Fig. 1 that, even though the facilitating effects of familiarization appear to be approaching an asymptotic level under Cond. A_{60}, performance under Cond. NA_0 was slightly better than under Cond. A_{60}. Irrespective of how proficient *S* becomes at pronouncing the stimulus units, it takes longer to pronounce than not to pronounce; it takes longer to say something than to say nothing.

Intercomparison of the various conditions via the critical difference technique (Linquist, 1953) revealed two significant ($P < .05$) differences, Cond. NA_0 vs. Cond. A_0 and Cond. NA_0 vs. Cond. NA_{60}. Inspection of acquisition as a function of trials for each of the conditions did not reveal evidence of interaction between treatments and trials.

In conclusion, it is apparent that PA instructions regarding *S*'s response to the stimulus term during the anticipation interval can determine whether SF facilitates or inhibits PA performance. We believe that these effects are attributable to covariation of the effective length of the anticipation interval with amount of familiarization. Furthermore, it may be expected that such factors as word length, pronunciability, and meaningfulness will also, depending on the length of the anticipation interval, interact with PA instructions and amount of familiarization.

References

Cieutat, V. J., Stockwell, F. E., & Noble, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m*, *m'*) in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 293–302.

Gannon, D. R., & Noble, C. E. Familiarization (*n*) as a stimulus factor in paired-associate verbal learning. *J. exp. Psychol.*, 1961, **62**, 14–23.

Linquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received September 14, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 550–551

SUPPLEMENTARY REPORT: TIME BETWEEN PAIRINGS AND SHORT-TERM RETENTION[1]

LLOYD R. PETERSON, KENNETH HILLNER, AND DOROTHY SALTZMAN

Indiana University

Peterson, Saltzman, Hillner, and Land (1962) found marked forgetting within 8 sec. after a single paired-associate presentation, when other presentations filled the interval. The present study investigates the effect of an 8-sec. interval similarly filled which is inserted between the first and second presentations of an individual pair later tested for retention.

Method.—The technique of the previous experiment was used. The first as well as the second pairing consisted of *S* reading aloud the stimulus and the response from the drum. Either 0 or 8 sec. separated the two pairings. The retention interval, measured from removal of the second presentation from the drum, was 2, 4, 8, or 16 sec. There were 42 *Ss* who were tested 16 times in each of the eight conditions. Stimuli were three- and four-letter words. The responses were indicated in the instructions to be the numbers 1–10. Twelve seconds rest separated 16 blocks of 31–41 exposures, save for a 1-min. rest between Blocks 8 and 9.

Results.—Table 1 shows that massed pairings resulted in superior recall at the 2- and 4-sec. retention intervals, while 8-sec. spacing was superior at the 8- and 16-sec. retention intervals. An analysis of variance found significance for both Spacing ($F = 10.72$, $df = 1/41$, $P < .01$) and Retention ($F = 71.80$,

TABLE 1

Proportions Correctly Recalled

Spacing Interval	First Half				Second Half			
	Retention Interval (Sec.)				Retention Interval (Sec.)			
	2	4	8	16	2	4	8	16
0 Sec.	.82	.69	.45	.48	.85	.74	.45	.46
8 Sec.	.81	.66	.56	.56	.82	.63	.60	.60

[1] This research was supported by Grant G 12917 from the National Science Foundation to Indiana University, a grant for which the senior author is principal investigator.

$df = 3/123$, $P < .01$). Their interaction was also significant ($F = 16.47$, $df = 3/123$, $P < .01$). The two halves of the session did not differ significantly ($F = .73$, $df = 1/41$, $P > .05$). After pooling data from the two halves of the session, individual t tests showed that differences between the two spacing conditions were significant at the 4-, 8-, and 16-sec. retention intervals ($t = 3.07$, 5.23, 4.84; $df = 41$, $P < .01$).

A paradox is presented by the finding that when an interval during which marked forgetting can be shown to occur is introduced between pairings, there is improvement in retention at long intervals. Underwood's (1961) explanation of distributed practice of lists does not seem appropriate here, since he concludes that distribution is superior to massing only when response learning is involved. Response learning was minimized by the instructions in the present study of spacing between individual pairings.

REFERENCES

PETERSON, L. R., SALTZMAN, D., HILLNER, K., & LAND, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

UNDERWOOD, B. J. Ten years of massed practice on distributed practice. *Psychol. Rev.*, 1961, **68**, 229–247.

(Received August 18, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 5, 551–552

SUPPLEMENTARY REPORT: YOKED COMPARISONS OF CLASSICAL AND AVOIDANCE EYELID CONDITIONING UNDER THREE UCS INTENSITIES[1]

I. GORMEZANO, JOHN W. MOORE, AND EDWARD DEAUX

Indiana University

Moore and Gormezano (1961) observed that classical conditioning when experimentally equated in terms of partial reinforcement pattern and number of UCS occurrences by a yoking procedure, was inferior to avoidance conditioning. The present investigation was conducted to determine the effects of UCS intensity on such yoked comparisons of classical and avoidance conditioning.

Method.—The general apparatus, procedure, and stimuli were the same as in the earlier experiment. The only variation was in the intensities of the UCS employed. Twenty *Ss* were assigned to each of the six cells of a 2×3 factorial design in which classical and avoidance conditions were made orthogonal to three UCS puffs of nitrogen which had intensities sufficient to support 40-, 80-, and 160-mm. columns of mercury. In addition 2 male *Ss* were lost because of apparatus failure. The two recording systems and sex were also made orthogonal to the classical-avoidance and UCS intensity dimensions.

Results.—The distributions of response latencies for all six groups in acquisition and extinction were recorded. The distributions (not shown) revealed that aside from the higher frequency of responses in the CR range for the avoidance groups, the only discernable difference was the tendency for the modal responses to decrease in latency, under both conditioning procedures, as UCS intensity increased.

Figure 1 presents the results of plotting percentage CRs for the six experimental groups in acquisition and extinction. The initial points on all acquisition curves are the mean percentage CRs on Trial 1 and the remaining points are for successive blocks of 10 trials. The extinction curves are plotted in 5-trial blocks. The figure indicates that the acquisition performance of the avoidance groups (A160, A80, A40), for each of the three UCS intensities, was superior to each of their respective yoked-classical groups (i.e., A160 vs. Y160, A80 vs. Y80, and A40 vs. Y40). Though Groups A160 and A80 follow essentially identical courses of acquisition, group Y160 was inferior to Y80. The extinction curves reveal that the higher level of responding of the avoidance groups in acquisition also persisted in extinction. As was observed in the previous study, CRs of the avoidance groups demonstrated steeper decay functions than the yoked-classical groups and are probably in large part due to their having started at higher levels of responding.

Split-plot analyses of variance (Snedecor,

[1] This research was supported by Grant G 16030 from the National Science Foundation. A report of this experiment was presented at the Psychonomic Society, New York, September 1961.

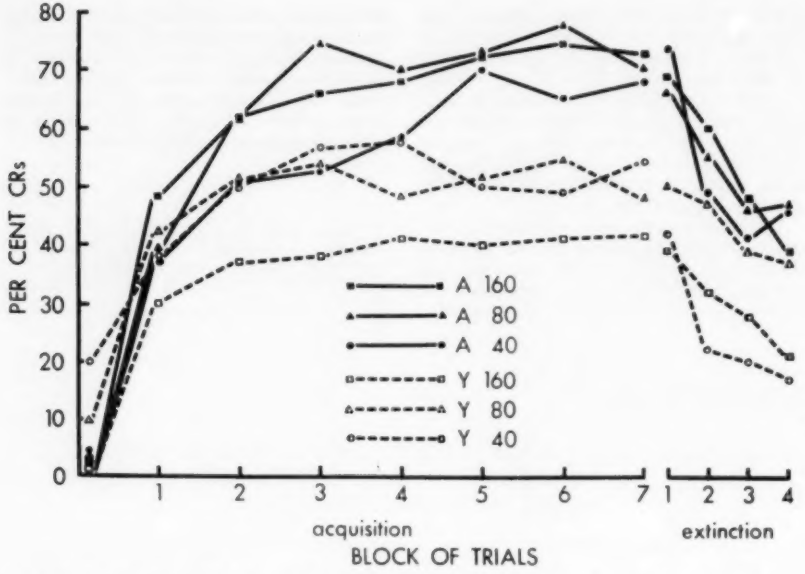

FIG. 1. The percentage CRs plotted in 10-trial blocks during acquisition and 5-trial blocks in extinction.

1959) on the arc-sine transform of individual percentage CRs for the 70 acquisition and 20 extinction trials revealed significant F values for the classical vs. avoidance comparisons in acquisition ($F = 27.15$, $df = 1/48$, $P < .005$) and extinction ($F = 22.55$, $df = 1/48$, $P < .005$). The UCS intensity dimension failed to produce significant differences in either acquisition or extinction. A significant F value was obtained in acquisition for the Classical-Avoidance × UCS Intensity interaction ($F = 14.39$, $df = 2/48$, $P < .005$), reflecting the fact that as UCS intensity increased, performance of the avoidance groups increased while performance of the yoked-classical groups decreased. This interaction appears to be simply a function of a negative correlation between the performance levels of the paired *Ss*. However, the significant interaction is heavily weighted by the low level of responding of Group Y160. If the performance of Group Y160 is not sampling error, its performance relative to Y80 suggests a Partial Reinforcement × UCS Intensity interaction. A significant F value was also obtained for the Classical-Avoidance × Sex interaction in extinction ($F = 5.02$, $df = 1/48$, $P < .05$); and examination of the mean percentage CRs revealed that under the avoidance procedure males gave a higher percentage of CRs whereas, under the yoked-classical procedure the females were superior. In the remaining sources of variation those of statistical significance were without psychological import (i.e., interactions involving the recording system source of variation).

REFERENCES

MOORE, J. W., & GORMEZANO, I. Yoked comparisons of instrumental and classical eyelid conditioning. *J. exp. Psychol.*, 1961, **62**, 552–559.

SNEDECOR, G. W. *Statistical methods*. (5th ed.) Ames, Iowa: Iowa State Coll. Press, 1959.

(Received September 7, 1962)

CONTENTS—*Continued*

On-Target Versus Off-Target Information and the Acquisition of Tracking Skill: A. C. Williams and G. E. Briggs 519

Resistance to Extinction after Varying Amounts of Discriminative or Nondiscriminative Instrumental Training: M. R. D'Amato, D. Schiff, and H. Jagoda 526

T Maze Reversal Learning after Several Different Overtraining Procedures: W. F. Hill, N. E. Spear, and K. N. Clayton 533

Monetary Incentive and Range of Payoffs as Determiners of Risk Taking: L. Katz ... 541

Spatial S-R Contiguity in Human Discrimination Learning: C. D. Standish and R. A. Champion 545

Supplementary Report: Stimulus Familiarization in Paired-Associate Learning: R. W. Schulz and I. F. Tucker 549

Supplementary Report: Time between Pairings and Short-Term Retention: L. R. Peterson, K. Hillner, and D. Saltzman 550

Supplementary Report: Yoked Comparisons of Classical and Avoidance Eyelid Conditioning under Three UCS Intensities: I. Gormezano, J. W. Moore, and E. Deaux 551

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich	Herausgeber für die Schweiz	Redaktion
PROF. DR. H. ROHRACHER, Wien	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Band IX/1

DIETER KLEBELSBERG, **Unterschiedliche Auswirkungen einer Blutalkoholkonzentration von 1, 0 Promille bei einzelnen psychischen Komponenten der Fahrtüchtigkeit**

ARNULF RÜSSEL, **Belastung und Ermüdung an unterschiedlichen Arbeitsplätzen**

KARL HEINZ SEIFERT, **Kompensatorische Leistungsmöglichkeiten Mindersinniger im Bereich der Haptik**

KONRAD STEININGER, **Aktuelle luftfahrtpsychologische Probleme**

LORÁND SZALAY, **Untersuchungen zur semantischen Struktur der Zeitwörter**

All articles are provided with summaries in English and French.
The journal appears quarterly and contains 160 pages per number.

Subscription is DM 64,—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
(Brentanoweg 10, Germany)